

# **Localisation of Virtual Sound Sources by Bilateral Cochlear Implant Users**

**Janani Cheryl Fernandez**

**School of Electrical Engineering**

Thesis submitted for examination for the degree of Master of  
Science in Technology.

Espoo 17.05.2017

**Thesis supervisor:**

Prof. Ville Pulkki

**Thesis advisor:**

PhD. Ville Sivonen

Author: Janani Cheryl Fernandez

Title: Localisation of Virtual Sound Sources by Bilateral Cochlear Implant Users

Date: 17.05.2017

Language: English

Number of pages: 6+45

Department of Signal Processing and Acoustics

Professorship: Acoustics and Audio Signal Processing (S-89)

Supervisor: Prof. Ville Pulkki

Advisor: PhD. Ville Sivonen

This thesis investigates the localisation of virtual sound sources by cochlear implant users. This research is intended to provide insight into the feasibility of utilising virtual sources to test their hearing abilities. In addition, it is hoped that the factors affecting the results can be determined by comparing the results to the patient demographic data. Listening tests were conducted with ten cochlear implant users and a control group of ten normal hearing subjects. Two methods of virtual sound source synthesis were applied: amplitude panning, via vector base amplitude panning, and time panning, via an empirical formula. A two alternative forced choice task was presented to the test subjects, and the results of the test were utilised to derive their minimum audible angle. It was found that amplitude panned sources can indeed be perceived by cochlear implant users and can, therefore, be utilised as a testing tool. Furthermore, it was found that the results were influenced by the age at which the subjects were diagnosed deaf and the time period between the diagnosis and implantation surgery.

Keywords: Cochlear Implant, Virtual Sound, Localisation, VBAP

## Preface

First, I would like to thank Dr. Ville Sivonen for advising me, finding volunteers for the listening tests, and organising the listening tests in the clinic.

Second, I thank Prof. Ville Pulkki for all his help and guidance, as well as giving me a desk to work on.

Third, all the researchers at the Acoustics lab for their feedback and support.

Last, but not least, I thank my friends and family: my parents, my sister, and especially my partner Leo McCormack, for their encouragement and for putting up with me while I wrote this thesis.

Otaniemi, 6.11.2017

Janani C. Fernandez

# Contents

<b>Abstract</b>	<b>ii</b>
<b>Preface</b>	<b>iii</b>
<b>Contents</b>	<b>iv</b>
<b>Symbols and Abbreviations</b>	<b>vi</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Hearing</b>	<b>3</b>
2.1 Peripheral Auditory System . . . . .	3
2.1.1 Outer Ear . . . . .	3
2.1.2 Middle Ear . . . . .	4
2.1.3 Inner Ear . . . . .	4
2.2 Spatial Hearing . . . . .	6
2.2.1 Monaural Cues . . . . .	7
2.2.2 Binaural Cues . . . . .	7
2.3 Summing Localisation and the Precedence Effect . . . . .	9
<b>3 Psychoacoustics</b>	<b>10</b>
3.1 Tasks . . . . .	10
3.2 Test Methods . . . . .	10
<b>4 Synthesis of Virtual Sound Sources</b>	<b>12</b>
4.1 Amplitude Panning . . . . .	12
4.1.1 Vector Base Amplitude Panning . . . . .	15
4.2 Time Panning . . . . .	16
<b>5 Technical Audiology</b>	<b>19</b>
5.1 Hearing Aids . . . . .	20
5.2 Cochlear Implants . . . . .	20
5.2.1 Binaural Hearing in CI Users . . . . .	23
<b>6 Methodology</b>	<b>27</b>
6.1 Subjects . . . . .	27
6.1.1 CI Group . . . . .	27
6.1.2 Control Group . . . . .	28
6.2 Test Set-up . . . . .	28
6.3 Stimuli . . . . .	29
6.4 Test Procedure . . . . .	30
6.5 Calculation of Results . . . . .	30
6.6 Limitations . . . . .	31
<b>7 Results and Discussion</b>	<b>32</b>

<b>8 Conclusion</b>	<b>39</b>
<b>References</b>	<b>40</b>
<b>A Appendix A</b>	<b>45</b>

# Symbols and Abbreviations

## Abbreviations

ASM	Automatic Sound Management
ASR	Active Stimulation Range
AGC	Automatic Gain Control
CA	CONTOUR ADVANCE
CI	Cochlear Implant
dB HL	Decibel Hearing Level
DRC	Dynamic Range Compressor
Hz	Hertz
HRTF	Head Related Transfer Function
ILD	Inter-aural Level Difference
ILDA	Inter-aural Level Difference Angle
ITD	Inter-aural Time Difference
ITDA	Inter-aural Time Difference Angle
kHz	Kilohertz
MAA	Minimum Audible Angle
NH	Normal Hearing
NIHL	Noise Induced Hearing Loss
Stand.	STANDARD
SS	SLIM STRAIGHT
SSD	Single Side Deaf
TAFC or 2AFC	Two Alternative Forced Choice
VBAP	Vector Base Amplitude Panning

# 1 Introduction

Out of the traditional five senses, hearing has numerous applications: including survival, communication, and entertainment. Therefore, partial or complete loss of hearing is considered to be a severe impairment to one's day-to-day life and quality of living. In 2012, it was estimated that 360 million people had some form of disabling hearing loss (WHO, 2012); therefore, it is important to understand hearing impairment and to improve the quality of life for persons afflicted with the condition. Before portable electronics were developed, this improvement was achieved through the use of mechanical devices to provide acoustic amplification, such as horns. During the 20th century, analogue hearing aids replaced such devices; they, in turn, have since been replaced by digital hearing aids and cochlear implants during the later half of the 20th century (Pulkki and Karjalainen, 2015). Hearing aids are external devices worn by the user, and typically function by amplifying and processing sound before it arrives at the tympanic membrane. On the other hand, cochlear implants (CIs) are devices that do not function as amplifiers, and instead bypass a large portion of the human hearing system; thus compensating for auditory system damage in someone with moderate to severe hearing loss (Clark, 2006).

There are two kinds of CI users: unilateral and bilateral, or in layman's terms, users with a cochlear implant in one ear and users with a cochlear implant in both ears. Initially, bilateral implantation was thought of as a technology upgrade, in which the latest CI model would be implanted in the second ear; nonetheless, studies have shown that this procedure was also effective in restoring binaural hearing ability to the user (Litovsky et al., 2004). Bilateral users can typically hear sounds better, have a better overall speech comprehension, require a lower signal to noise (SNR) ratio to comprehend speech in noisy environments, and have better localisation abilities than unilateral CI users (Litovsky et al., 2006). Due to such findings, bilateral implantation has been performed from the 1990s onwards (Van Hoesel and Clark, 1997; Van Hoesel et al., 1993). In Finland, bilateral cochlear implantation has become clinical routine for congenitally deaf children and, depending on individual assessments, for adults as well (Välimaa et al., n.d.; Härkönen et al., 2015).

While these technologies do much to aid persons with hearing loss, they have not been perfected, nor have their effects been fully understood. CI users have varying rates of adapting to the devices, and the causes for this variance have not been fully understood at the time of writing this thesis, partially due to the heterogeneity of the population of CI users (Tobey et al., 2000; Clark, 2006). In bilateral CI users, the ability to localise sound also varies across users (Litovsky et al., 2006, 2009). In a recent study, Dorman et al. (2016) found a large variation in localisation results for bilateral CI users, while the variation for normal hearing participants was notably smaller; however, they did not find a conclusive reason for the variation. Additionally, it should be noted that testing the localisation abilities with real sound sources can be a difficult task, as the results are limited by the set-up utilised for the tests.

This thesis investigates CI users' localisation abilities with regards to amplitude and time panned sound sources. Additionally, it attempts to discover a correlation between CI users' patient demographic data, such as age and time of implantation,

and the results of the localisation tests. It is hoped that this investigation will provide some insight into user adaptability to their devices, as well as confirm CI users can hear simulated sources, thus providing researchers another tool, that is not as limited by the loudspeaker set-up, to investigate their abilities.

This thesis is organised as follows: the first chapter of this thesis examines hearing, with the human auditory system explained in detail. The second chapter provides some information on psychoacoustics and the methods utilised in the listening test. The next chapter explains the virtual sound sources and the algorithms utilised to produce these sources for the test. The fifth chapter elaborates on technical audiology, particularly with respect to CIs. Following this is the methodology that elaborates in detail the listening test set-up and the methods utilised therein. The results and discussion chapter contains the findings of the test, and the examination of these results.



## 2 Hearing

Human hearing is a physical and neurological process performed by the auditory system, which consists primarily of the auditory cortex located in the brain and the peripheral auditory system. The typical frequency range of human hearing, in someone with normal hearing, is 20 Hz to 20 kHz. This range decreases with age, particularly in the higher frequencies (Moller, 2012). Human hearing is particularly sensitive in the frequency range 1 kHz to 4 kHz, which is the range that contains the most information in spoken communication (Gold et al., 2011).

In this chapter, the human auditory system is explained in detail. In addition, spatial hearing and the factors that contribute to it are discussed, with an emphasis on those cues that are relevant to the thesis.

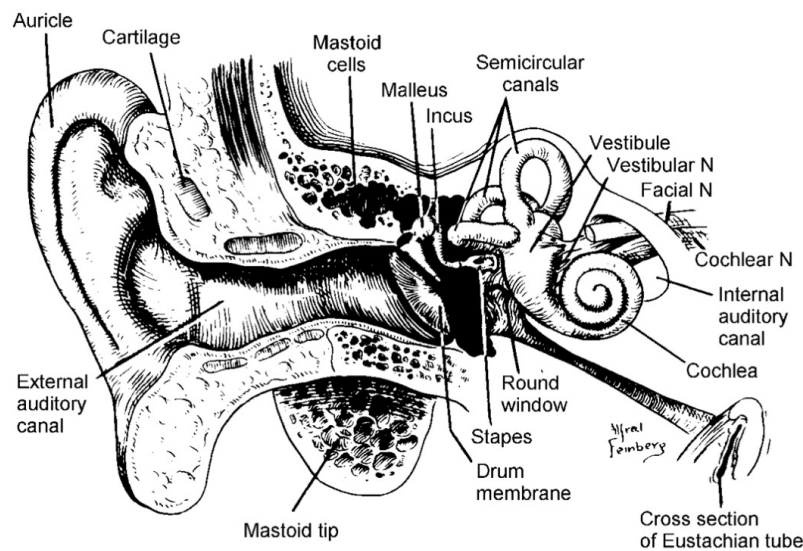


Figure 1: Diagram of the human auditory system, adopted from Davis and Silverman (1970).

### 2.1 Peripheral Auditory System

The left and right ear form the peripheral audio system and are the primary sensors of sound in the majority of mammals. The ear is a complex organ that can be divided into three sections: the outer ear, the middle ear, and the inner ear.

#### 2.1.1 Outer Ear

The outer ear consists of the *pinna*, *auditory canal*, and *tympanic membrane*. The pinna is composed of cartilage and contributes to the direction dependent filtering of the torso; therefore, it aids in localisation of sound sources, particularly for frequencies

above 4 kHz (Pulkki and Karjalainen, 2015). Furthermore, it occasionally attenuates background noise caused by wind (Moore, 2012). The auditory canal is a passageway or cavity through which sound waves travel to the *tympanic membrane* or, as it is more commonly known, the *ear drum*. These sound waves cause the tympanic membrane to vibrate, and are then transmitted through the membrane to the middle ear.

### 2.1.2 Middle Ear

The function of the middle ear is to perform impedance matching between the outer ear and the fluid-filled inner ear. It also provides some protection from loud sounds by contracting the bones in such a manner that the transmission of vibrations is attenuated, essentially acting as a biological dynamic range compressor (DRC) (Moore, 2012).

The tympanic membrane is connected to the middle ear, which consists of three small bones called the *malleus*, *incus*, and *stapes*. They are also known as the *hammer*, *anvil*, and *stirrup* and are collectively called the ossicles, which are the smallest bones in the human body. The vibrations of the tympanic membrane are coupled to the *malleus*, then to the *incus*, and finally to the *stapes*; thus causing each to vibrate in turn. The stapes is connected to an opening, termed as the oval window which leads to the inner ear. The vibrations from the stapes pass into the inner ear through the oval window.

### 2.1.3 Inner Ear

The inner ear has a labyrinthine shape, which can be divided into two sections: the bony labyrinth and the membranous labyrinth (Marieb and Hoehn, 2007). The latter is enclosed within the bony labyrinth, which is a maze-like series of bone channels in the temporal bone. The fluid in the bony labyrinth is called the perilymph, and the fluid in the membranous labyrinth is called the endolymph (Marieb and Hoehn, 2007).

The bony labyrinth consists of the cochlea, the vestibule, and the semicircular canals. The cochlea is connected to the oval window and is a spiral shaped organ with walls of bone; in humans, the cochlea has approximately 2.5 turns (Moller, 2012). The end of the cochlea connected to the oval window is called the base, while the other end is called the apex. Two membranes, the basilar membrane and Reissner's Membrane, divide the cochlea lengthwise (Pulkki and Karjalainen, 2015). At the basal end, the basilar membrane is smaller in width and mass, as well as being stiffer; contrarily, at the apical end it has the opposite properties, as it is wider, more massive, and supple (Marieb and Hoehn, 2007).

The fluid filled canal above the basilar membrane is called the *scala vestibuli*, while the one beneath the basilar membrane is referred to as the *scala tympani*; the fluid in these sections is called the perilymph and they are connected by a opening, called the *helicotrema*, at the apical end of the basilar membrane (Marieb and Hoehn, 2007). A third fluid canal, called the scala media, is formed by Reissner's membrane and the basilar membrane (Pulkki and Karjalainen, 2015). It is the scala media, also

known as the cochlear duct, which contains the endolymph (Marieb and Hoehn, 2007). The Organ of Corti, which is the sensory organ, resides on the basilar membrane within the cochlear duct (Moore, 2012).

The vibrations from the base of stapes are connected to the perilymph through the oval window, causing the perilymph in the cochlear to vibrate. The perilymph is nigh incompressible and the function of another opening in the cochlea, the round window, is to move in opposite phase to the oval window; thus allowing for movement of the fluid (Alberti, 2001).

Due to these varying physical properties of the basilar membrane, each point along its length is sensitive to its own characteristic frequency; therefore, the vibrations of the perilymph will cause sections of various frequencies contained in the audio signal to vibrate (Pulkki and Karjalainen, 2015). The closer the section of the basilar membrane is to the oval window, the higher the frequency to which it is sensitive. Hence, the basal section of the cochlear is sensitive to high frequencies, while low frequencies are detected at the apical end (Pulkki and Karjalainen, 2015).

The endolymph within the cochlear duct has a high concentration of  $K^+$  ions in comparison to the perilymph (Marieb and Hoehn, 2007). This leads to a potential difference across the basilar membrane as there is a difference in cation concentration in the cochlear duct and the scala tympani on either side of the membrane (Pulkki and Karjalainen, 2015). The Organ of Corti on the basilar membrane contains approximately 16,000 hair cells placed along the membrane. These hair cells are sensory organs which are sensitive to  $K^+$  ions (Moller, 2012). The hair cells are arranged in four rows, and the function of the outer rows of hair cells differs from the function of the inner hair cells. The movement of the hair cells caused by vibrations in the basilar membrane will open the pathway for the  $K^+$  cations to interact with stereocilia located at the tips of the hair cells (Moore, 2012). This interaction leads to electrical signals being transmitted through the auditory nerve fibres, which are connected to the hair cells (Pulkki and Karjalainen, 2015). As the hair cells are arranged along the basilar membrane, the information transmitted to the auditory cortex contains information about the frequency, envelope, and amplitude of the sound waves arriving at the ear (Moller, 2012).

However, computational models of the above mentioned functioning of the cochlea results in a frequency selectivity that is not as sharp as those measured clinically from auditory nerve responses (Moller, 2012). It was theorised that this was due to the existence of a neural "second filter"; however, this was not confirmed (Moller, 2012; Marieb and Hoehn, 2007). Furthermore, mechanical measurements of the vibration pattern of the basilar membrane were performed on cochleas obtained from cadavers. It was found that after death, the frequency selectivity of the basilar membrane deteriorated; therefore, it was concluded that there was some form of active functioning of the auditory system, which contributes to the frequency selection (Moller, 2012).

## 2.2 Spatial Hearing

The localisation of sound is the ascertaining of the position or origin of a detected sound (Pulkki and Karjalainen, 2015). The benefit of having a pair of ears is the existence of binaural cues which the brain can utilise to localise sound by performing real-time processing of the two audio signals received by the ears. The localisation of sound sources and reflections provide an understanding of the spatial environment the listener is in and aids in navigation. In general, humans are better at localising sound in left-right directions rather than top-down directions. Furthermore, while the localising of the sound in the majority of the spherical coordinates is considerably accurate, humans struggle with estimating the distance of the sound from their current position (Moller, 2012). Localisation is performed using a combinations of cues, which may be monaural or binaural in nature (Blauert, 1997).

Before elaborating further on these cues, it is important to describe the three dimensional coordinate system utilised in their description. In general, a spherical

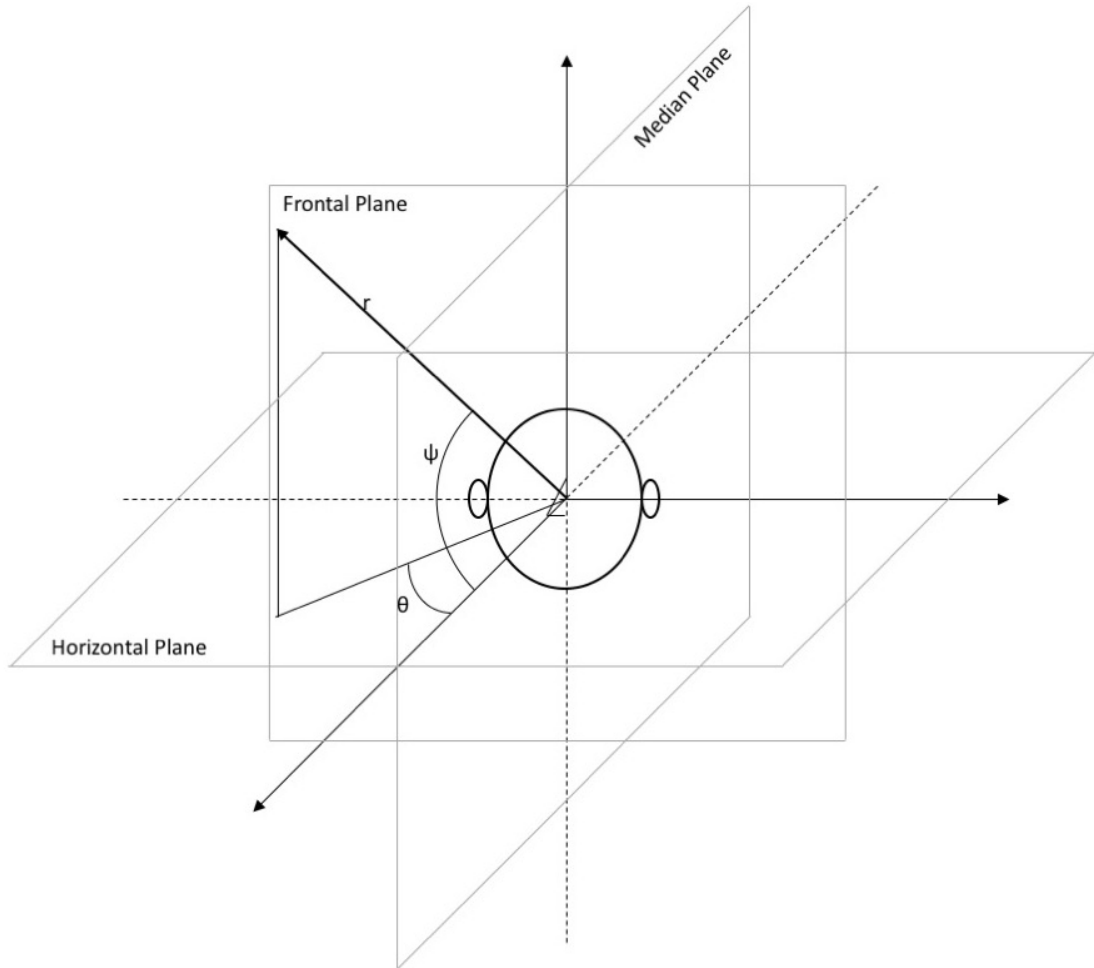


Figure 2: Diagram of spherical coordinate system.

coordinate system is a more preferable format to describe the system as it is natural and intuitive for this purpose. This system is composed of three planes: the median plane, which lies vertically and divides the head in halves by passing between the two eye sockets and through the nose and mouth; the horizontal plane, which is perpendicular to the median plane, and horizontally divides the head in two by passing through both the ears; and the frontal plane, which is perpendicular to both the median and horizontal plane. The position of a point in the spherical system is described by three parameters: the azimuth angle  $\theta$ , elevation angle  $\psi$ , and radius distance  $r$ .

The Minimum Audible Angle (MAA) is the slightest angular change in the location of a sound source that is discernible to a listener. It is a discrimination threshold and is dependent on the stimuli and the location of the sound source relative to the listener, for example, the MAA for pure tones for an azimuth angle directly in front of the listener is approximately  $1^\circ$  (Blauert, 1997).

Additionally, the torso and pinna contribute to the Head Related Transfer Function (HRTF), which is the pressure measured in anechoic conditions at a point in the ear canal divided by the pressure measured in anechoic conditions with the torso absent, at the point where the centre of the head was located for the first measurement (Møller et al., 1995). The HRTF can be utilised to predict the response to a sound source from a point in space (Pulkki and Karjalainen, 2015). Both the binaural and monaural cues are collectively encapsulated in an HRTF. The phase characteristics of HRTFs contain ITDs, while the ILD is the frequency-dependent magnitude differences between two HRTF spectra. The monaural cues are derived from the transfer function between the source signal and the signal arriving at the ear drum for each ear.

### 2.2.1 Monaural Cues

The scattering caused by the torso and pinna are direction and frequency dependent, leading to spectral cues which are monaural in nature. Monaural cues may be defined as cues which are derived from the audio signal of a single ear.

The pinna has many complex folds and cavities, and is asymmetrical along all axes. Additionally, it is highly individualistic; therefore, techniques developed for one listener may not translate well for another listener. The reflections due to the pinna aid in localisation along the median and frontal planes, and is considered to be important at frequencies above 4 kHz (Blauert, 1997).

### 2.2.2 Binaural Cues

Binaural cues are derived from the differences between the auditory information received by the auditory nerve fibres at the two cochleas. These cues are broadly divided into two categories: inter-aural time differences (ITDs) and inter-aural level differences (ILDs) (Pulkki and Karjalainen, 2015). The contribution of these cues towards localisation varies across frequencies (Pulkki, 2002).

## ITD

Inter-aural Time Differences (ITDs) are the phase differences between the audio signals received by the two auditory channels (Pulkki and Karjalainen, 2015). These differences are primarily caused due to the differences in the path of sound waves to the two ears; thus, the arrival time of the wavefront will be different between the two tympanic membranes. For example, sound waves from a source to the left of the head will have to travel a longer path to reach the right ear than the left.

It has been found that a person, said to possess normal hearing, is capable of utilising ITDs for localisation for frequencies up to 1.5 kHz; however, for higher frequencies, the wavelength is short, which may render the phase differences ambiguous (Buell and Hafter, 1991; Pulkki, 2002). At frequencies above 1.5 kHz, ITDs impart information primarily on envelope differences and are, therefore, considered not to be a strong localisation cue at these frequencies (Moore, 2012; van de Par and Kohlrausch, 1997; Pulkki, 2002).

## ILD

The Inter-aural Level Difference (ILD) is the level differences between the sound received at the two ears, which is influenced partly by the sound waves propagating through the medium for different distances and also the interaction of the sound waves with the listener's head. However, this effect is largely dependent on the dimensions of the human head, which cast an acoustical shadow on the contralateral ear signal, which can cause a reduction in level as great as 15 dB for frequencies above 2 kHz (Shaw, 1974; Duda and Martens, 1998). The sound pressure level at the ear closest to the sound source is larger in amplitude than the other ear, which biases the localisation of the source towards the direction of the ear closest to it.

Yost and Dye Jr (1988) discovered that the ILD can be detected for the majority of audible frequencies. Additionally the threshold for ILD detection is considerably low. On the other hand, scattering effects are prominent when the proportion of the size of the head to the wavelength is sufficiently large for the head to affect the wave. Hence, the relevance of ILD in localisation is limited to frequencies in which the level differences are sufficiently large, which are mainly the frequencies above 1 kHz (Shaw, 1974). Consequently, the ILD is a strong localisation cue at high frequencies and a weak cue at low frequencies.

## Dynamic Cues

These cues are created by movements of the head, causing changes in the binaural cues. It is particularly useful in resolving sound source position uncertainty in the cone of confusion, which is a perceptually ambiguous region where the ITDs and ILDs are identical (Pulkki and Karjalainen, 2015).

## 2.3 Summing Localisation and the Precedence Effect

A reflection may be defined as copy of the original sound source that is smaller in amplitude, has a different direction of arrival, and is delayed (Litovsky et al., 1999). If the delay between the original sound and the reflections is below the echo threshold, they are fused together and perceived as a single sound; the echo threshold is typically between 30 ms and 40 ms (Pulkki and Karjalainen, 2015). Additionally, the *precedence effect* occurs, in which the location of the combined sound is perceived to be in the direction of the original sound. Should the delay be 1 ms or less, the *summing localisation* effect occurs whereby the original sound and the reflections are combined and contribute to the localisation of the sound, which will then be localised to a position in between the original sound and the reflection (Blauert, 1997).

### 3 Psychoacoustics

Psychophysics is the science in which perceptual processes are studied by altering one or several characteristics of a stimulus, typically one that can be objectively measured (Smith, 1997). Psychoacoustics is the branch of psychophysics that deals with auditory perception. For approximately a century psychoacoustics has sought to explore auditory perceptions, and ultimately find a correlation between psychoacoustic phenomena and human physiology.

As this thesis utilises psychoacoustic experiments to examine the localisation abilities of CI users, this chapter provides an overview of psychoacoustic tasks and methods. The psychoacoustical experiments are largely non-invasive and relatively effortless in comparison with other fields, such as neurology. Typically, the aim of these experiments is to form the psychometric function. The function models or represents the relationship between the changes in the stimuli and the sensation and perceptions produced by the stimuli. This function is formed based on the responses of the test subject to a particular task given to them. It should be noted that the majority of psychoacoustic tasks and methods are common to both psychoacoustics and psychophysics as a whole.

#### 3.1 Tasks

There are various types of tasks in psychoacoustics, some of which are listed and explained below:

**Threshold:** the aim of this task is to find the absolute threshold of a stimulus, with threshold defined as the magnitude of the stimulus that is required for the test subject to perceive said stimulus for a set fraction of the time (Pulkki and Karjalainen, 2015). Typically, the fraction is set at 50%.

**Discrimination:** the purpose of the task is to discover the difference threshold, i.e. the smallest difference between two stimuli required for the subject to perceive the difference for a set proportion of time (Pulkki and Karjalainen, 2015).

**Detection:** the aim of this task is for the subject to report if they detect the stimulus, typically within an interval.

**Forced choice:** in these experiments the subject is asked a question, then given a number of options and must choose one based on the question. These tasks may be utilised to determine the threshold or they may be scored. One subset of forced choice tasks is the two alternative forced choice (TAFC or 2AFC) tasks in which two options are presented to the subject, with no indifference option present.

#### 3.2 Test Methods

These tasks are presented to subjects utilising various testing methods, some of which are elaborated upon below:

**Method of constant stimuli:** in this method, the test utilises a significant number of stimuli with values around the assumed value of the threshold. A test is then carried out for each value; the task for the test depends upon the threshold to



be determined, and is preferably a multiple interval or alternative forced choice task to mitigate bias. The psychoacoustic function is then formed based on the results of the test. It may be possible for the function to not approach 0, in which case a 2AFC test may be conducted.

**Method of limits:** in this method, the subject is presented with a detection task with the level of the stimulus altered either automatically or by the experimenter. The results of this test are then utilised to determine the threshold.

**Method of adjustment:** this is more of a task than a method. The subject is tasked with varying an attribute of the stimuli until it matches a given reference value.

**Method of tracking:** this is a method whereby the subject is able to affect the direction of change of the attribute being examined. One commonly known test of this type is Békésy audiometry, in which the subject is given a detection task to press a button when they perceive a sound; the level is the frequency swept tone is decreased upon pressing the button and the average of the results is termed as the absolute hearing threshold of the subject.

**Direct Scaling:** this is a common method which presents one or more reference samples alongside the stimuli; the reference samples are typically termed as *anchors*. An example of this method is the popular MUSHRA test (Pulkki and Karjalainen, 2015).

**Adaptive staircase:** in this method the level of the studied attribute is changed depending on whether the subject answered the given forced choice test correctly. If the subject answered correctly, the level is changed to make the test more difficult; on the other hand, if the answer was incorrect, the level difference renders the test easier. The value of the studied attribute plotted against the number of tests resembles a staircase, hence the name of this method. Ideally, the function will converge to the sought threshold after an adequate number of reversals. An average of the studied attribute value over the last few rehearsals is taken as the threshold value. The change in the levels is based on a given step size, and tests are often designed to begin with relatively large step sizes, which are then reduced to smaller step sizes nearer the assumed time of convergence.

Levitt (1971) formed several implementations of the staircase technique technique, including a technique that he named the '2-down 1-up' track. It consisted of conducting two tests, then checking the results of the test. If both tests were correct, the level of the attribute value changed in order to make the test harder; however, if even one test was wrong the attribute value was altered to make the test easier. Then two tests were conducted again, with the previous steps repeated until convergence was achieved. Levitt then determined the threshold by taking the average of the attribute value for the last reversals, and he proved that this was the threshold for correct responses 70.7% of the time.

## 4 Synthesis of Virtual Sound Sources

A virtual sound source is a sound source that is perceived by the listener to be at a position in which there is no real sound source. They are typically synthesised via a suitable reproduction of a monophonic audio signal. Typically, the sole attribute of the virtual source being manipulated is the direction, however, in some instances its width or distance from the listener are also controlled.

In this chapter, two methods of virtual source synthesis are discussed: amplitude panning and time panning. Furthermore, as this thesis deals with localisation of these virtual sources, the localisation cues present in the virtual sources, synthesised by these two methods, are examined.

The production of virtual sound sources requires a loudspeaker set-up consisting of two or more loudspeakers. There exists a listening position within a loudspeaker set-up in which the method is most effective; this position is known as the *sweet-spot* and is commonly located at the position where all loudspeakers are equidistant to the listener.

If the listening position is located at a point that is not equidistant from the loudspeakers, compensation must be applied i.e. appropriate gains and delays must be applied to the audio signals fed to the loudspeakers to compensate for the unequal distances. Magnitude compensation is also applied in set-ups which do not utilise identical loudspeakers.

The most common loudspeaker configuration is the stereophonic 2-channel set-up, which utilises 2 loudspeakers, positioned such that there is a  $60^\circ$  angle between them, with the position of the listener at the vertex. Increasing the number of loudspeakers generally leads to an increase in the perceived spacial accuracy of the sound source.

### 4.1 Amplitude Panning

This method of virtual sound source synthesis is achieved by altering the gain or amplitude of the audio signal given to each loudspeaker in the set-up. Mathematically, amplitude panning can be represented by the equation

$$x_i(t) = g_i x(t), i = 1, \dots, N \quad (1)$$

where  $x(t)$  is the input signal for time instance  $t$ ,  $x_i(t)$  is the signal given to  $i$ th loudspeaker,  $g_i$  is the gain for the  $i$ th loudspeaker, and  $N$  is the number of loudspeakers (Pulkki and Karjalainen, 2015). Each panning technique utilises a panning law to estimate the direction of arrival based on the gains applied to the loudspeakers. Furthermore, to prevent changes in the loudness level the sum of the gains must be normalised (Pulkki and Karjalainen, 2015). Mathematically

$$\sum_{n=1}^N g_n^p = 1. \quad (2)$$

In the case of reverberant rooms, the value  $p = 2$  was found to be suitable (Pulkki and Karjalainen, 2015). The normalisation method may be different, depending on the characteristics of the room the listener is within (Moore, 1990).

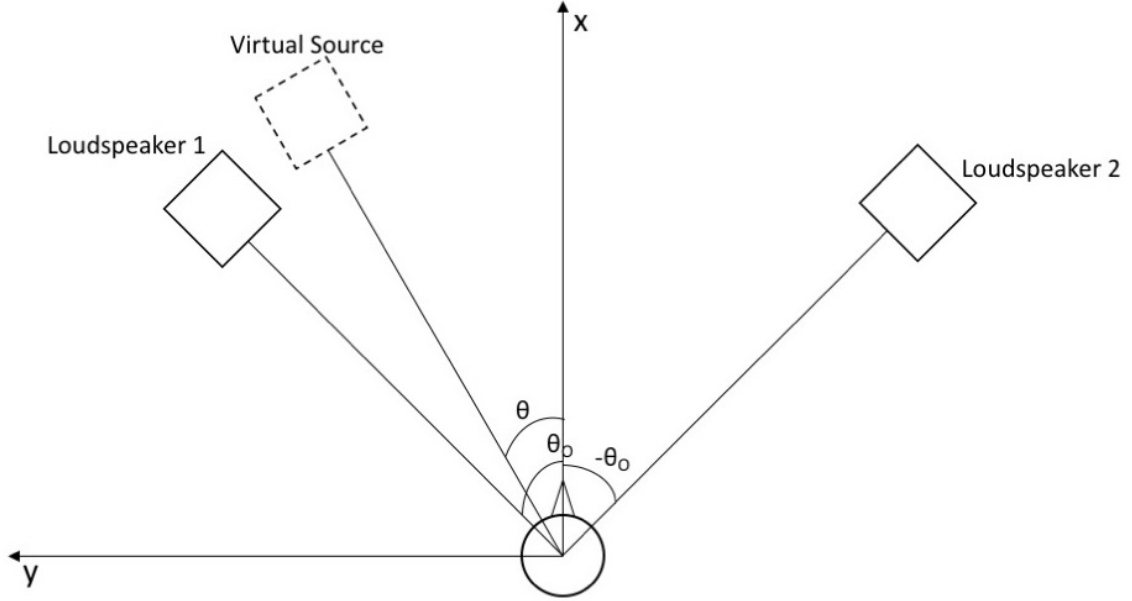


Figure 3: Virtual sound source production utilising amplitude panning.

As stated before, a stereophonic set-up is the most commonly utilised set-up and the sine and tangent laws may be applied to perform amplitude panning in such a set up. The sine law is represented mathematically as

$$\frac{\sin \theta}{\sin \theta_0} = \frac{g_1 - g_2}{g_1 + g_2}, \quad (3)$$

where  $\theta$  is the estimated direction of arrival of the perceived sound source,  $0^\circ < \theta < 90^\circ$ ,  $\theta_0$  is the azimuth angle between the loudspeakers divided by 2, and  $g_1, g_2$  are the gains applied to the audio signals fed to loudspeakers 1 and 2, respectively, where  $g_1, g_2 \in [0, 1]$ .

The sine law is limited as it is only applicable when the listener remains in the forward facing position. The tangent law, however, does not share this limitation and the variations caused by the movement of the head of the listener is not perceptible as they are negligibly small (Pulkki, 1997). The law is represented mathematically as

$$\frac{\tan \theta}{\tan \theta_0} = \frac{g_1 - g_2}{g_1 + g_2}, \quad (4)$$

where  $\theta$  is the estimated direction of arrival of the perceived sound source and  $0^\circ < \theta < 90^\circ$ ,  $\theta_0$  is the azimuth angle between the loudspeakers divided by 2, and  $g_1, g_2$  are the gains applied to the audio signals fed to loudspeakers 1 and 2, respectively, where  $g_1, g_2 \in [0, 1]$ .

At low frequencies, it takes advantage of summing localisation; the amplitude difference between the loudspeakers is interpreted as a phase difference as the

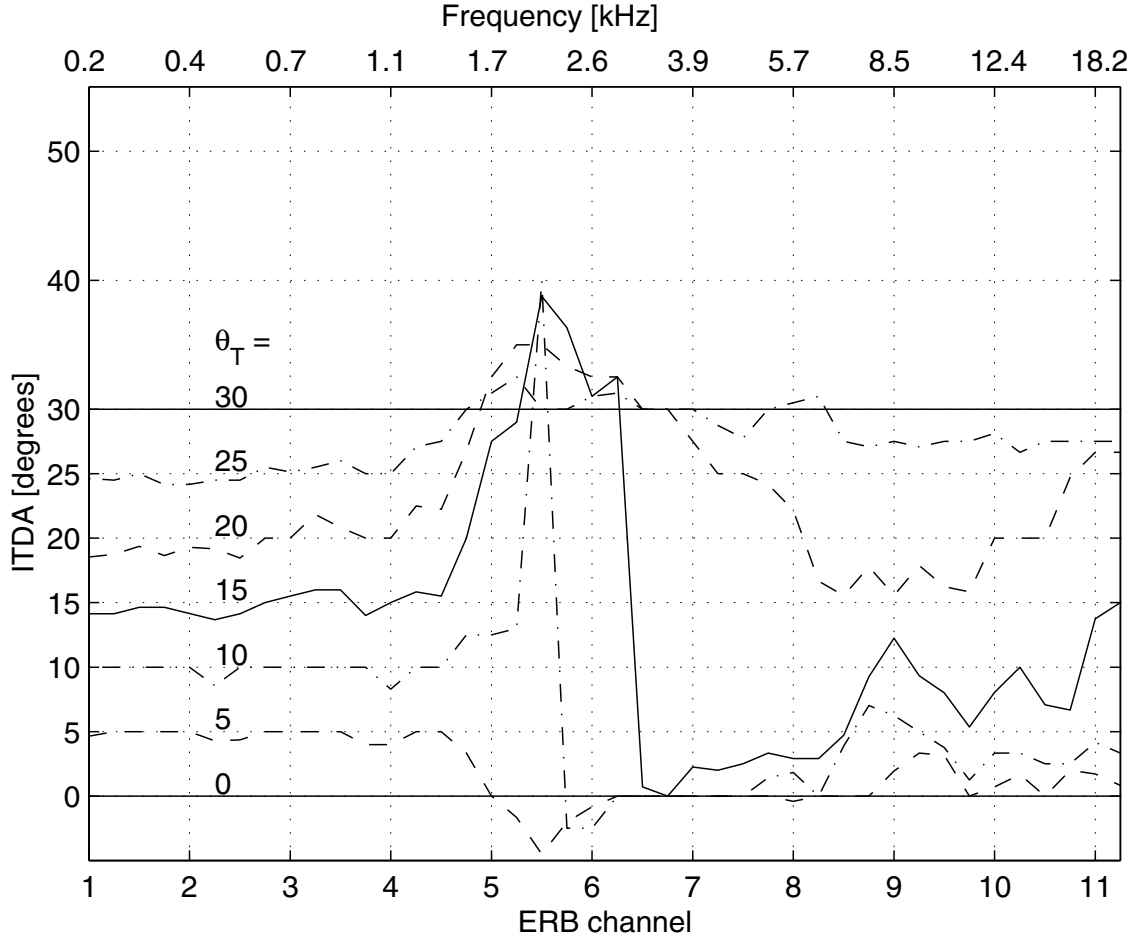


Figure 4: Simulated ITD cues of amplitude panned virtual sources plotted as a function of panning angle  $\theta_T$  and frequency; adopted from Pulkki et al. (1999).

shadowing of the head is not prominent at these frequencies (Pulkki et al., 1999). At high frequencies, the head shadowing effect and the reflections off the torso and pinna are more prominent, which potentially distorts the ILD cue; nonetheless, with the right conditions, the amplitude difference of the loudspeakers is interpreted as the ILD cue (Pulkki et al., 1999).

Figures 4 and 5 plot the localisation cues of broadband amplitude panned virtual sources for various panning angles and frequencies, in which ITDA is the inter-aural time difference angle, and ILDA is the inter-aural level difference angle, the angular extrapolations of ITD and ILD respectively. It can be seen that for low frequencies, the ITD cues are flat and close in value to the panning angle; contrarily, above 1.1 kHz the ITDA is unstable and rises sharply before dropping to near zero values. In comparison, the ILDA is more stable at high frequencies than the ITDA. As the ILD cues are relied upon more than ITD cues at high frequencies, the deviation in ITDA cues is not perceptible; however, the slight instability in the ILD values could result in the virtual sound source being perceived as spatially spread. Furthermore, there is a region, from 1.1 kHz to 2.6 kHz, in which both the ITDA and the ILDA

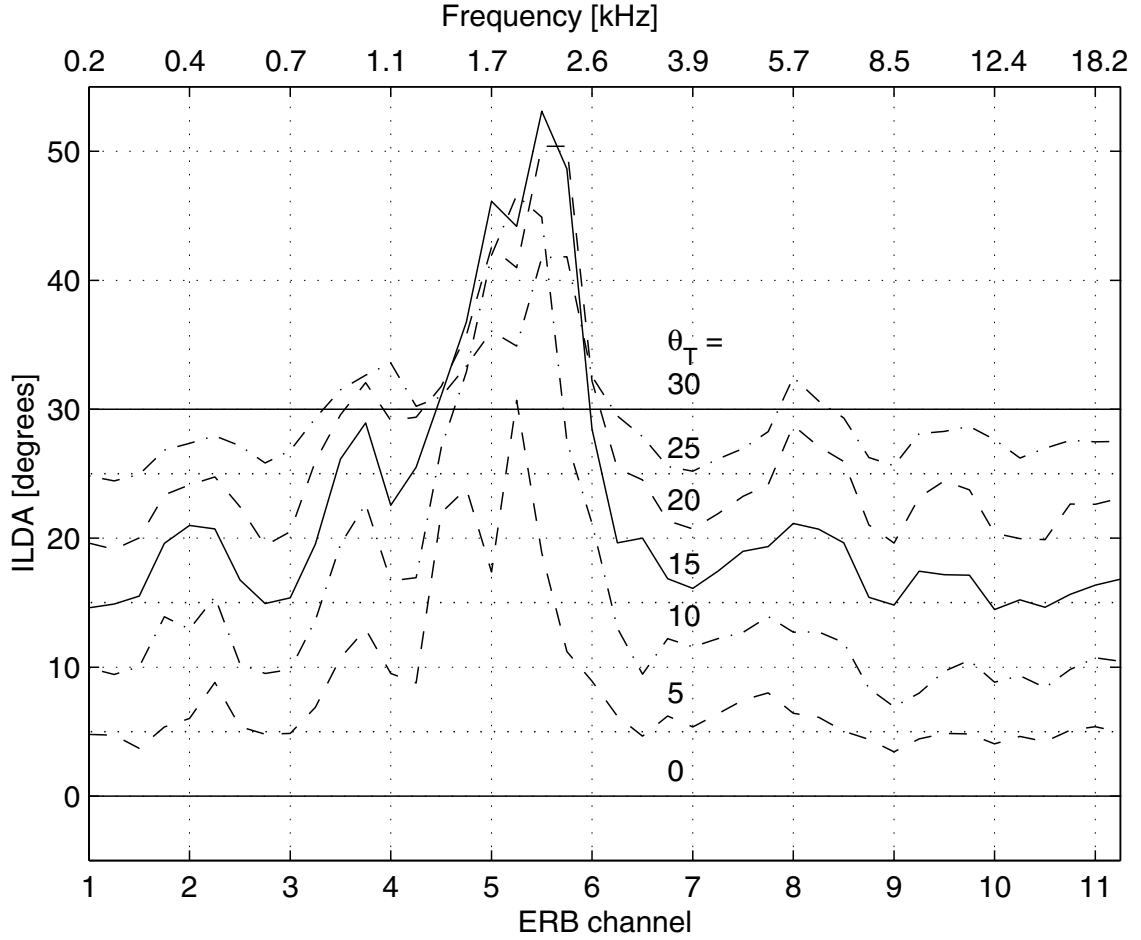


Figure 5: Simulated ILD cues of amplitude panned virtual sources plotted as a function of panning angle  $\theta_T$  and frequency; adopted from Pulkki et al. (1999).

are unstable; therefore, localisation of amplitude panned sounds of frequencies in this region could be ambiguous.

#### 4.1.1 Vector Base Amplitude Panning

The previous panning laws discussed were trigonometric formulations of the loudspeaker set up, while VBAP is a vector base formulation of the loudspeaker set-up. For this method of panning, the set-up is defined by two vectors  $l_1$  and  $l_2$  are of unit magnitude or length and directions that point in the direction of the loudspeaker positions.

In a stereophonic loudspeaker set-up,  $l_1 = [l_{11} \ l_{12}]^T$  and  $l_2 = [l_{21} \ l_{22}]^T$ . The direction of the virtual source can be defined by a unit vector  $p = [p_1 \ p_2]^T$ . As  $p$  is a vector, it can be defined in terms of  $l_1$  and  $l_2$  as

$$p = g_1 l_1 + g_2 l_2 \quad (5)$$

where  $g_1$  and  $g_2$  are the loudspeaker gains. Replacing  $l_1$  and  $l_2$  in equation (5), then

rearranging the equation, results in the following (Pulkki, 1997):

$$[g_1 \ g_2] = p^T L_{12}^{-1}, \quad (6)$$

where  $L_{12} = [l_1 l_2]^T$ . Note that the above equation cannot be solved if the inverse of  $L_{12}$  does not exist, which occurs if the elevation angle is  $\psi \neq 0^\circ$  and  $\psi \neq 90^\circ$ . Furthermore, the normalisation is performed utilising the equation:

$$g = \frac{\sqrt{C}g}{\sqrt{g_1^2 + g_2^2}} \quad (7)$$

in which  $C$  is a constant that can be interpreted as the volume control for the amplitude panned source (Pulkki, 1997).

This principle can be expanded and applied to set-ups with more than two loudspeakers, as well as 3-dimensional set-ups. In the case of 2-dimensional multi-loudspeaker set-ups, the loudspeakers pairs are formed then, depending on the required direction of arrival, an appropriate pair is fed the scaled audio signals. In the case of a 3-dimensional set-up, triplets are formed in the place of pairs and three gains are calculated as opposed to two gains.

## 4.2 Time Panning

Time panning is achieved by applying a constant delay to one loudspeaker, resulting in the listener perceiving the position of the virtual source to be closer to the non-delayed loudspeaker; thus taking advantage of the precedence effect (Pulkki and Karjalainen, 2015).

This method of panning is similar to spaced microphone recording techniques, in which microphones are positioned with gaps between them, as the space between microphones in such a technique produces time delays between the recordings, thus resulting in a similar spatial perception as time panning. A spaced microphone recording arrangement of two microphones placed 17 cm apart is a close approximation to a stereophonic time panning set-up; therefore, the localisation cues of the microphones can be considered to be the equivalent to the cues in a time panning set-up. Figure 6 displays the simulated cues of such a microphone recording. At low frequencies the ILDA oscillates significantly; while at high frequencies has near  $0^\circ$  for all directions of arrival. In comparison to the ILDA, the ITDA is more stable and a closer match to the true direction of arrival. At 200 Hz it has near zero values, then the ITDA rises rapidly before stabilising at relatively larger values at higher frequencies for all directions. Furthermore, the ITDA and ILDA cues tend to suggest opposing directions of arrival. The strange behaviour of the cues indicate localisation would be dependent on frequency, as the cues vary with frequency (Pulkki, 2002). Additionally, with short click-like sounds only the ITDA is quite salient; while, for continuous sounds, both the ITDA and the ILDA are salient. This leads to spatial spreading in the perception of non click-like time panned sounds, while the click-like sounds are typically perceived as points in space when time panned (Pulkki, 2002).

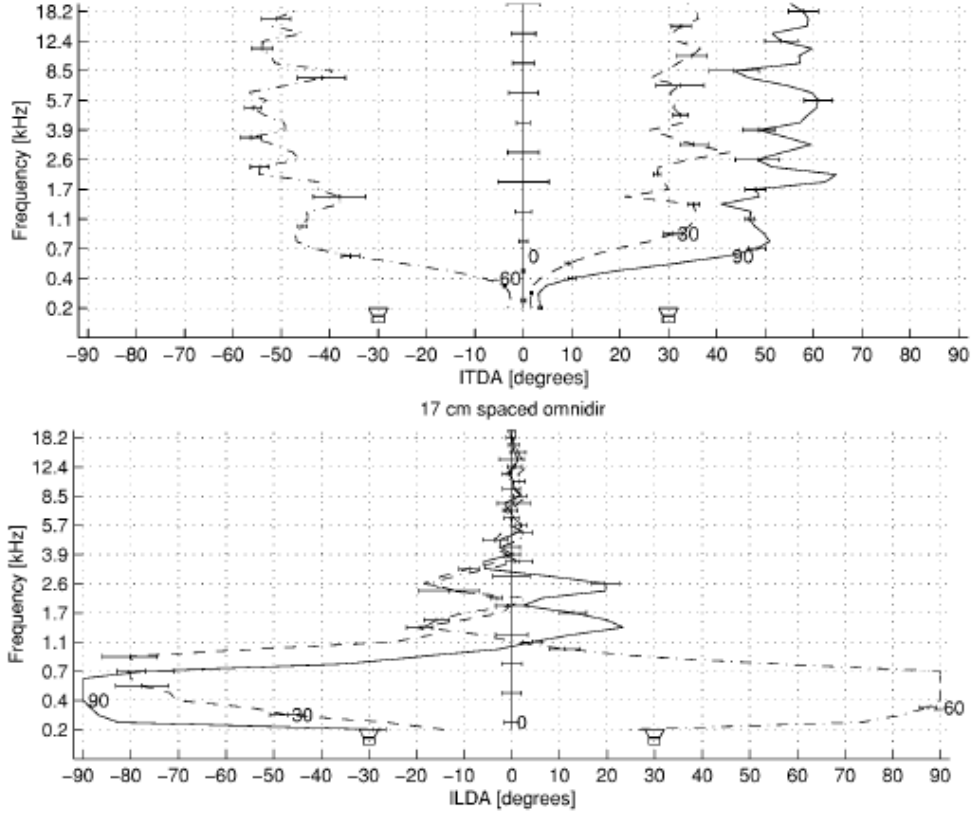


Figure 6: Simulated localisation cues of a spaced omnidirectional microphone arrangement with two microphones placed 17 cm apart. The cues are calculated utilising six individual HRTF sets, then the mean and standard deviation determined. The whiskers of the box plots denote a 25% standard deviation; adopted from Pulkki (2002).

Thus, at low frequencies, the time or phase delays between the signal emitted by the two loudspeakers is interpreted as level differences, i.e. ILDs; while at high frequencies, the delays are interpreted as ITDs. This interpretation of cues is dependent on the listening position, since change in the listening position alters the delays, as well as the sound signal itself (Pulkki and Karjalainen, 2015; Lipshitz, 1986).

There have been attempts to utilise both amplitude and time panning to synthesise virtual sources, such as the method proposed by Lee (2017). Lee developed an empirical formula through listening tests with expert listeners. The formula calculates the delays and gains required to implement a specified hybrid of amplitude and time panning; it may also be utilised solely for amplitude panning or solely for time panning with the following equation

$$ICTD = 0.025\left(\frac{30\theta}{\theta_0}\right); \quad (8)$$

where ICTD is the delay, in samples, that is applied to the delayed channel,  $\theta$  is the

panning angle, and  $\theta_0$  is the loudspeaker base angle (Lee, 2017).



## 5 Technical Audiology

This chapter presents a broad overview of hearing aids and cochlear implants. Additionally, the various studies of the localisation abilities of CI users are discussed in detail. The results of these studies will later be compared to the results obtained in this thesis.

Given the number of people that suffer from hearing loss, the prevention of and the compensation for the condition has been investigated for quite some time. The most basic of devices for enhancing hearing is the mechanical trumpet. Currently, the preferred devices for the hearing impaired population are hearing aids and cochlear implants (Pulkki and Karjalainen, 2015).

Hearing loss can be categorised utilising various methods. One method of categorisation is by severity, which is typically three categories: mild, moderate, and severe (Davis, 1989). This categorisation defines the loss in dB HL, which is decibels Hearing Level. dB HL is measured via an audiogram and is relative to the smallest or quietest sound a person with normal hearing is capable of perceiving; thus, a person with 5 dB HL hearing threshold can hear sounds only when they are 5 dB louder than the level at which a person with normal hearing can perceive them. Mild hearing loss is defined as hearing loss of 25 dB HL or greater, moderate hearing loss is a loss of 45 dB HL or greater, and severe hearing loss is a loss of 65 dB HL or greater (Davis, 1989).

Another method of hearing loss categorisation is by cause, which can be broadly divided into three categories: conductive hearing loss, sensorineural hearing loss, and mixed hearing loss (Moller, 2012). Conductive hearing loss is caused by a failure to conduct sound waves to the ear canal. It typically presents itself as a reduction in loudness. The causes for the conduction failure include illness, obstruction or damage to the ear canal, damage to the ossicles, and unequal pressure before and after the tympanic membrane (Moller, 2012). Conductive hearing loss may be temporary or healed through medical treatment; however, in some cases it is permanent.

Sensori-neural hearing loss can be further divided into two categories: Neural and Sensory; however, recent evidence indicates that these two categories may not be as separate as previously believed (Moller, 2012). Neural hearing loss is primarily caused by disorders in the brain and auditory nerve, i.e. problems which originate in the nervous system. It can be caused by age, damage to the auditory nerve or cortex, or changes in the neural system. Sensory hearing loss is rooted in problems of the cochlea. It is commonly caused by pathologies of the cochlear hair cells, such as damage to hair cells or distension of the basilar membrane. Causes for sensory hearing loss include congenital disorders, Noise Induced Hearing Loss (NIHL), and age (Moller, 2012). Neural hearing loss can be temporary, but it is often permanent.

Mixed hearing loss is a combination of sensorineural and conductive hearing loss. It may be due to a number of causes, such as NIHL, age, and illness.

## 5.1 Hearing Aids

Hearing aids are utilised to aid persons with mild to moderate hearing loss (Pulkki and Karjalainen, 2015). Their primary function is to amplify and process sound before it enters the auditory canal and they are effective in combating most types of permanent conductive hearing loss and some forms of sensori-neural hearing loss (Moller, 2012). The amplification of the sound signal causes vibrations of a sufficient amplitude that they are perceived by damaged hair cells or are able to overcome obstructions.

A hearing aid typically consists of a single unit per ear that comprises of a microphone, processing unit, and loudspeaker (Pulkki and Karjalainen, 2015). The microphone records the audio signal, while the loudspeaker plays the amplified signal into the auditory canal. As hearing loss is rarely of the same level across frequency bands, the processing unit in a hearing aid ensures the appropriate amplification in various frequency bands for the individual user. The devices may be behind the ear, in-the-ear, or in-the-canal; users choose a style depending on their comfort level, price, appearance, and budget.

## 5.2 Cochlear Implants

Cochlear implants are typically utilised in cases of moderate to severe hearing loss (Moller, 2012). Their function is to replace the damaged sections of the auditory system; they bypass the peripheral auditory system and directly stimulate the auditory nerve (Clark, 2006). They are effective in cases of sensory hearing loss, as well as conductive hearing loss (Moore, 2012).

A cochlear implant is composed of two units per ear, one unit is implanted into the mastoid bone of the skull and the other unit is typically worn behind the ear, although some CI users may have the second unit in other positions such as the back of the head (Clark, 2006). The implanted unit has minute wires which conclude in electrodes that are positioned along the basilar membrane; these electrodes provide the stimulus to the auditory nerve in the form of electrical pulses in which the audio signal is encoded; thus, a cochlear implant can compensate for any damage or problem in the auditory system except damage, problems, and disorders in the auditory cortex and auditory nerve (Moller, 2012). The second unit of the cochlear implant has a microphone within it, as well as a digital signal processor. The function of the unit is to transform the audio signal received by the microphone into the electric signal that is transmitted through the skin to the implanted unit, then on to the auditory nerve through the electrodes (Clark, 2006). Their frequency range is typically from 100 Hz to 8 kHz.

All models of cochlear implants mimic the behaviour of the auditory system by utilising bandpass filters to divide the spectrum of the input audio signal into several channels, then feeding the frequency bands into a channel that feed one or two electrodes; they also perform some form of compression, typically with an automatic gain control (AGC) or a dynamic range compressor DRC or both (Clark, 2006). Furthermore, some processors also compute information such as formant frequencies,

fundamental frequencies, and temporal characteristics to enhance the performance of the devices (Moller, 2012). Modern devices have between 12 and 22 intracochlear channels; studies have proven that a minimum of four to five channels are required to provide an adequate degree of speech intelligibility (Fishman et al., 1997; Loizou, 1999).

The directivity of CI devices can be omnidirectional, fixed directional, or adaptive. Cochlear and MED-EL have the various settings for the directionality that include: omnidirectional, adaptive, or Natural. An adaptive directionality is described as one which adapts to the situation and changes the microphone directivity accordingly; while the Natural directivity is said to be a directivity pattern that mimics that of a normal ear. There can be several programmable modes with various combinations of microphone directionality and sensitivity, noise reduction algorithm settings, and gain settings that the user can switch between, depending on the situation.

After being fitted with an implant, users require time for their auditory cortex to adapt to the device and to learn to interpret the signals. This adjustment period varies in time for individuals. Välimaa et al. (2011) discovered statistically significant improvements in vowel recognition took place for upto 18 months in best performers and 12 months for poor performers. Oh et al. (2003) concluded that in children, speech perception improved over the entire four year period of observation, while the average results of the adult participants did not show improvement after two years.

The length of the cochlea is ear dependent, thus the active length of the electrode array must be optimised for each ear and each person. The length of the electrode array over which the electrode contacts are placed is known as the active stimulation range (ASR). There are several electrode arrays commercially available with varying ASRs, and typical arrays aim to cover well over one turn, with no channels protruding outside the cochlea. However, occasionally, part of the ASR protrudes outside the cochlea, leading to undesirable effects such as pain, or facial tics. In such instances, the channels that cause these effects will be deactivated, reducing the number of active channels for that device.

In this study, the subjects utilised MED-EL and Cochlear brand devices for their cochlear implants. MED-EL devices have electrode arrays consisting of 12 channels. There are four different series of arrays: FLEX series, FORM series, CLASSIC series, and the ABI array. The CI users who participated in the listening tests for this thesis utilised arrays from the CLASSIC and the FLEX series.

- The CLASSIC series has been part of MED-EL's catalogue for more than 20 years, and consists of wave-shaped wires with 24 electrode contacts placed in pairs along the electrode array (Med-El, 2017b). The electrode arrays in this series are: the STANDARD, an electrode array that is 31.5 mm in total length, an ASR of 26.4 mm, basal diameter of 1.3 mm, apical diameter 0.5 mm; the MEDIUM, an electrode array of total length 24 mm, ASR 20.9 mm, basal diameter 0.8 mm, and apical diameter 0.5 mm; and finally, the COMPRESSED, an electrode array 15 mm in total length, an ASR of 12.1 mm, basal diameter 0.7 mm, and apical diameter 0.5 mm. The COMPRESSED was specially designed for malformed cochleas and partial ossification (Med-El, 2017b).

- The FLEX series has flexible, tapering, and soft electrode arrays of varying lengths (Med-El, 2017b). The FLEXSOFT is 31.5 mm in total length, an ASR of 26.4 mm, a diameter of 1.3 mm at the basal end and 0.5x0.4 mm at the apical end; the FLEX28, FLEX24, and FLEX20 are 28 mm, 24 mm, and 20 mm total length respectively and an ASR of 23.1 mm, 20.9 mm, and 15.4 mm respectively. Furthermore, they all have basal diameters of 0.8 mm and apical diameters of 0.5x0.4 mm. All electrode in this series have 19 electrode contacts, 14 of which are placed in pairs (Med-El, 2017b).

Similar to MED-EL, Cochlear produces several electrode arrays, of these, CI subjects in this thesis utilised the CONTOUR ADVANCE and SLIM STRAIGHT electrode arrays.

- The CONTOUR ADVANCE is a pre-curved tapered electrode array with 22 electrode contacts and ASR of 15 mm, a length of 24 mm, a basal diameter of 0.8 mm, and an apical diameter of 0.5 mm (Cochlear, 2017a).
- The SLIM STRAIGHT has 22 electrode contacts and ASR of 20 mm, a basal diameter of 0.6 mm, and an apical diameter of 0.3 mm (Cochlear, 2017b).

Each brand utilises its own unique sound processors. The CI users with MED-EL devices had Sonnet, Rondo, and Opus2 processors in their devices, while the CI users with Cochlear devices had CP810 and CP910 processors in their devices.

- The SONNET processor from MED-EL has several features which include dual microphones, Automatic Sound Management (ASM) 2.0, wind noise reduction algorithms, and automatic volume control with a dual loop AGC (Med-El, 2017d). ASM 2.0 is a strategy to ensure CIs have the best performance in the environment around them; its key features include natural and adaptive microphone directionality, a compression algorithm, programmable and an automatic volume control, and a programmable wind noise reduction algorithm (Med-El, 2017a). The pre-programmed settings of the ASM 2.0 can be alternated by the user with a remote controller. Furthermore, the SONNET CIs have an acoustic stimulation range up to 2 kHz (Med-El, 2017d).
- The RONDO processor from MED-EL is a single unit processor that utilises ASM, an older generation sound management strategy that does not have wind noise reduction or natural and adaptive microphone directivity; however, it is programmable and the pre-programmed settings can be selected utilising a remote controller (Med-El, 2017c). The ASM utilises a dual stage AGC, automatic and programmable volume control, and a compressor (Med-El, 2017a). As the Rondo is a single unit processor, the device is conveniently concealed in the user's hair; however, the microphone on the device is pointed towards the posterior of the skull.
- The OPUS 2 processor from MED-EL has a remote controller that allows the user to select the pre-programmed settings which best suit the situation.

- The CP910 sound processor from Cochlear is also programmable, and the user can either alternate between pre-programmed settings utilising a remote controller, or allow the device to adapt to the environment; the settings are labelled as "quiet", "speech", "speech in noise", "wind", and "music". The units are equipped with two microphones and is therefore able to perform beamforming which makes the various settings possible.
- The CP810 sound processor from Cochlear can store a number of pre-programmed settings, which the user can choose from using a remote controller. The settings include "everyday", "music", and "noise".

### 5.2.1 Binaural Hearing in CI Users

Localisation and spatial awareness is dependent on the auditory cortex receiving auditory information from both ears. In cases of hearing impairment typically both ears are stricken with hearing loss, however, the CIs are focused on improving speech intelligibility, with localisation sidelined. Non-speech frequencies are attenuated while speech frequencies are occasionally enhanced, which may alter localisation cues.

In order to ensure that information from both ears is delivered to the auditory cortex, unilateral CI users may have a hearing aid fitted to the other ear; however, their localisation abilities are only marginally better than that of a monaural listener and it is theorised that this is caused by the additional time incurred before the auditory nerve receives the information, which is caused by the processing time of the hearing aid (Dorman et al., 2016). Better results are achieved with binaural CI users, and several studies investigating their localisation abilities discovered that CI users regain and develop them after the implantation of the second CI, (van Hoesel and Tyler, 2003; Laszig et al., 2004; Nopp et al., 2004). Several papers comparing unilateral and bilateral CI users have found that bilateral CI users show notably higher localisation results and speech intelligibility than unilateral CI users (Dunn et al., 2008; Dorman et al., 2016). Additionally, their spatial awareness increases with time as CI users learn to associate the acoustic space to their surroundings (Zheng et al., 2015).

Studies exploring the spatial hearing capabilities of CI users discovered they are particularly sensitive to ILDs and their ITD cue is severely diminished (Seeber and Fastl, 2004; van Hoesel and Tyler, 2003; Lawson et al., 1998). However, they are to some extent sensitive to envelope ITDs as some devices encode data on both ILDs and envelope ITDs; therefore in CI users, envelope ITDs play a larger than usual role in localisation (Seeber and Fastl, 2008). Dorman et al. (2014) studied the ILDs in CI users and found their localisation error depended on the magnitude of the ILDs, as after being processed by the CI device the ILDs were notably diminished in magnitude. For persons with normal hearing, ITDs play a major role in localisation of low and mid frequency sound sources while ILDs play a major role in localisation of high frequency sound sources. In contrast to this, CI users display an inverted reliance on localisation cues and rely heavily on ILDs for localisation of sound sources, while the contribution of envelope ITDs play in minor role in localisation (Seeber and Fastl, 2008).

Additionally, the information received from reflections within the pinna are not present in the audio signal due to the position of the microphone; thus, disrupting the reflection and scattering effect of the human body, which results in significantly altered HRTFs. CI users who suffered hearing loss later in life must adapt to the altered HRTFs in addition to training their brains to interpret the information received by the CIs as well as adapt to poor frequency resolution provided by the device. On the other hand, CI users who were stricken with hearing loss from birth must learn to localise sound for the first time, in addition to adapting to interpreting sound and speech in general.

Due to the heavy reliance on one localisation cue, CI users do not have the ability to utilise redundancy to aid localisation in situations that require more information or to deal with uncertainties, such as speech in noisy situations or in situations with more than one sound source (Seeber and Fastl, 2008). Therefore, it may be difficult to correlate results of localisation tests with the results of standard speech in noise tests. Another problem may be the differing nature of the tests, as speech perception can be done monaurally. Several studies have reported no connection between localisation and speech intelligibility scores, both in quiet and in noise (Kerber and Seeber, 2012; Dunn et al., 2008; Tyler et al., 2006). On the other hand, one study has reported that there is a correlation between speech intelligibility scores in quiet situations and localisation (Litovsky et al., 2009). The results of this study are to some degree disputed as the correlation may exist due the general performance of CI users who have adapted well to their devices or have devices that are better fitted to the user (Kerber and Seeber, 2012).

Various studies have analysed CI users' localisation thresholds in the horizontal plane. Grantham et al. (2007) found the root mean square (RMS) error in localising noise stimuli ranges between  $8.1^\circ$  to  $43.4^\circ$  degrees, with a mean of  $24.1^\circ$ . In the same tests, CI users fared better in the localisation of speech stimuli, with a mean RMS error of  $21.5^\circ$ ; this indicates CI users find speech easier to localise. Another study found the mean RMS error in localisation to be  $29^\circ$  (Neuman et al., 2007). In a more recent study, Dorman et al. (2016) compared the localisation of CI users, normal hearing, and hearing aid users. Figure 7 is the plot of the results from this study. It was discovered that the mean RMS error was  $29^\circ$  for bilateral CI participants, while the young normal hearing participants had a mean RMS error of  $6^\circ$ , with older participants having a mean RMS error of  $5.4^\circ$ . It was noted that the best performing CI subjects had scores slightly higher than the 95th percentile score for the young normal hearing subjects. Furthermore, the variation in RMS errors for the normal hearing participants is significantly smaller than the variation for the other participant groups. The reasons behind the large variation in localisation abilities for the hearing impaired population has not been conclusively established, at the time of writing this thesis.

Litovsky et al. (2006) assessed localisation in 13 children with bilateral CIs and discovered that their MAA in the horizontal plane ranged between  $5^\circ$  and  $26^\circ$ ; comparatively, children without hearing loss and of similar age as the subjects in the test have MAAs between  $1^\circ$  and  $2^\circ$  (Litovsky et al., 2006). This considerable difference in localisation ability may be linked to the age at which the CI users

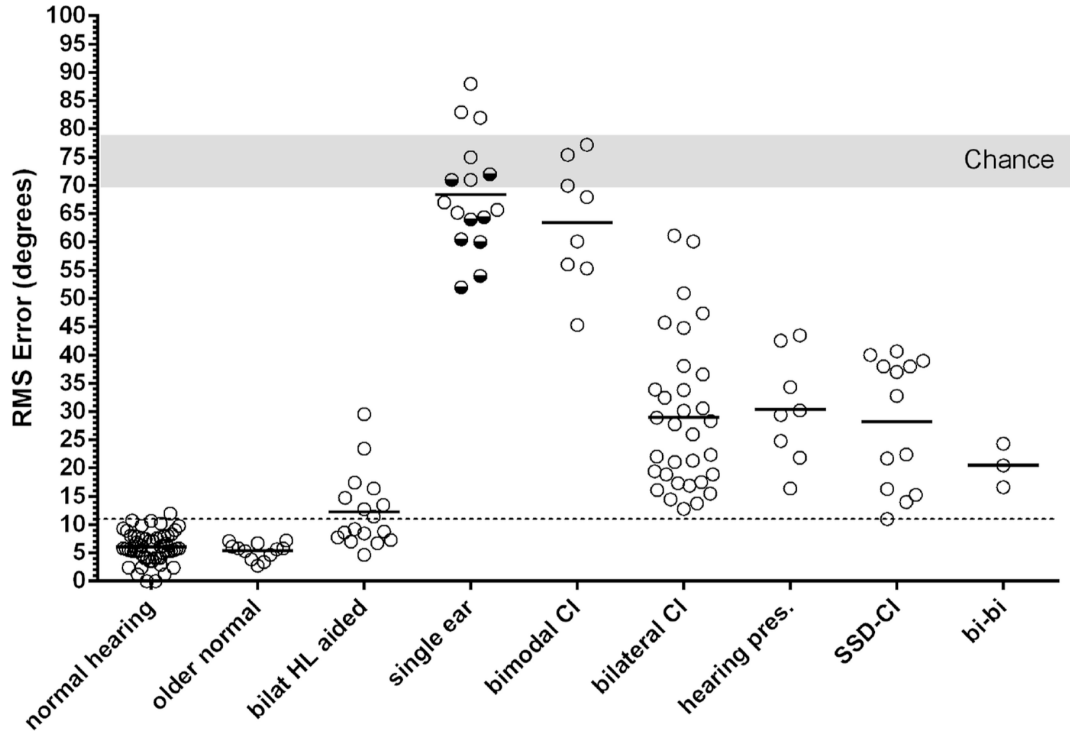


Figure 7: Localisation results, subjects separated into groups labelled young normal hearing, old normal hearing, bilateral hearing aids (bilat HL aided), single CI or single normal hearing ear (single ear), bimodal CI, bilateral CI, bilateral hearing preservation subjects, single side deaf (SSD) and one CI subjects, and bilateral CI with bilateral hearing preservation (bi-bi); adopted from Dorman et al. (2016).

received their implants, as the participants in Litovsky's tests that were two years old and had received their implants before the age of one displayed MAAs that were within normal limits (Grieco-Calub and Litovsky, 2012). Additionally, the study noted that the subjects improved their score when they received training. Another study which evaluated the localisation of CI users in the horizontal plane concluded their MAA ranged between  $3^\circ$  and  $8^\circ$ ; however, only five bilateral CI subjects participated in this study (Senn et al., 2005).

There has been very little research of virtual sound source localisation by CI users. A case study, Seeber and Fastl (2004), created virtual sources by manipulating the ILDs and ITDs, and determined that, in the horizontal plane, the CI subjects' localisation errors for the virtual sources were similar to their errors for real sources. Other studies focused on virtual sound production through headphones utilising HRTFs. Majdak et al. (2011) measured individual HRTFs of his subjects and tested their localisation abilities in both the horizontal and the median plane. They concluded that CI users can identify the correct left or right hemifield in the horizontal plane, as well as locate the position of the virtual sound source within the hemifield with reasonable precision; however, in the median plane they can only identify the correct front or back hemifield. Another study, Sechler et al. (2017), also utilised

HRTFs to create virtual sound sources. In addition, they used virtual reality headsets for their listening tests. The RMS error calculated for CI subjects was  $23.81^\circ$ , while normal hearing subjects had an RMS error of  $12.4^\circ$ . Additionally, they noted that the RMS error of CI subjects varies with the azimuth angle, unlike with the NH subjects.



## 6 Methodology

Little research has been conducted on the subject of virtual sound source localisation among the hearing impaired population; therefore, the methods applied were psychoacoustic methods adapted to suit the situation at hand. This thesis aims to investigate the utility of virtual sound sources in the testing of CI users. Hence, listening tests were conducted to determine the MAA of the CI subjects in comparison to a control group of NH subjects. This chapter elaborates on the methodologies applied in this thesis. The test subjects and their relevant information are detailed. Additionally, the test set-up, stimuli, test procedure, and the method of calculation of results are discussed.

### 6.1 Subjects

The participants in the tests are divided into two groups: the CI Group, which consists of CI users who volunteered for the experiments; and the control group or NH group, which consists of the normal hearing subjects.

#### 6.1.1 CI Group

A total of ten CI users volunteered for the tests. Each participant had at least two years of experience with both implants; hence, they are assumed to have had ample time to adapt to their devices.

Sub	Sound Processor		Directivity		Electrode Array		Active	
	Left	Right	Left	Right	Left	Right	Left	Right
CI1	Sonnet	Opus2	Omni	Omni	Flex28	Flex24	9	9
CI2	Rondo	Rondo	Omni	Omni	Flex28	Flex28	12	12
CI3	Opus2	Sonnet	Omni	Natural	Flex28	Flex28	11	11
CI4	Sonnet	Opus2	Natural	Omni	Flex28	Flex24	12	10
CI5	Opus2	Sonnet	Omni	Adaptive	Stand.	Flex28	12	12
CI6	Sonnet	Opus2	Omni	Omni	Stand.	Flex28	12	12
CI7	CP910	CP910	Everyday	Everyday	CA	CA	22	22
CI8	CP810	CP910	Everyday	Everyday	CA	CA	22	22
CI9	Opus2	Sonnet	Omni	Omni	Stand.	Flex28	11	12
CI10	CP810	CP910	Everyday	Everyday	CA	SS	22	22

Table 1: CI Group: Device Data

Table 1 consists of the relevant data about their devices. The subjects are labelled as CI1 through to CI10. All subjects possessing a MED-EL device except for subjects CI7, CI8, and CI10. The subjects with Cochlear devices had their noise reduction algorithms disabled for the duration of the test. The table lists the sound processors in each ear, the directionality of their programmes for each ear, the electrode array per ear, and the number of channels active in each array. Certain terms have been abbreviated to conserve space. These terms are: 'Stand.', in place of the STANDARD

electrode array; 'CA', in place of the CONTOUR ADVANCE; and 'SS', in place of the SLIM STRAIGHT.

Sub	Age	Implant Side		Time Deafened		Time Between Implants	Bilateral Exp.	Age Diagnosed
		CI1	CI2	Left	Right			
CI1	41.4	Right	Left	6.2	8.5	2.3	2.4	30.5
CI2	37.8	Left	Right	27.9	29.0	1.1	2.6	6.1
CI3	37.4	Left	Right	1.1	2.6	1.5	1.9	32.9
CI4	37.9	Right	Left	12.9	14.1	1.2	2.6	21.3
CI5	61.4	Left	Right	10.4	24.3	13.9	2.7	34.3
CI6	55.2	Right	Left	27.3	40.2	12.9	2.7	12.2
CI7	43.8	Left	Right	30.5	35.8	5.3	2.7	5.3
CI8	42.1	Left	Right	32.7	37.6	4.9	2.8	1.7
CI9	50.5	Left	Right	43.0	47.8	4.8	2.6	0
CI10	21.7	Left	Right	4.4	10.3	5.9	2.2	9.2

Table 2: CI Group: Patient Data.

Table 2 lists the relevant patient data of the CI subjects. All subjects, with the exception of subject CI6, were female. It may also be noted that subject CI4 was the pilot subject. In Table 2, the implant side column lists which ear contained the first implant, and which ear had the second implant. The following column, time deafened, is the time, in years, between diagnoses of deafness and the date of implantation. The time between implants column lists the time between the dates of the first and second surgery in years, while the bilateral experience column is the time, in years, between the date of the second surgery and the date of the listening test. Finally, the age diagnosed is the age, in years, of the subject at time of diagnoses of deafness. The average age of this group was 40.9 years, the average time between surgeries was 5.4 years, and the average bilateral experience was 2.5 years.

### 6.1.2 Control Group

Table 3 lists the data of the ten NH subjects. The subjects are labelled as NH1 through to NH10, and their age (in years) listed along with their gender. Their average age was 39.9 years. None of the subjects reported problems with their hearing, however no pure tone audiometry tests were performed.

## 6.2 Test Set-up

The listening tests were conducted in a specially designed listening room known as a sound field room that was 2.54 m in length, 2.64 m in width, and 2.10 m in height. The room contained eight loudspeakers positioned in a circular arrangement around a chair, two of which were utilised in the experiment. They were positioned at a

Sub	Age	Gender
NH1	46.0	Female
NH2	50.0	Female
NH3	43.0	Female
NH4	37.0	Female
NH5	36.0	Male
NH6	33.0	Female
NH7	38.0	Female
NH8	30.0	Male
NH9	51.0	Female
NH10	41.0	Female

Table 3: NH Group: Patient Data.

base angle of 45 degrees and a distance of 0.98 m from the chair. The room had a door and window. The diagram of this listening arrangement is shown in Figure 8.

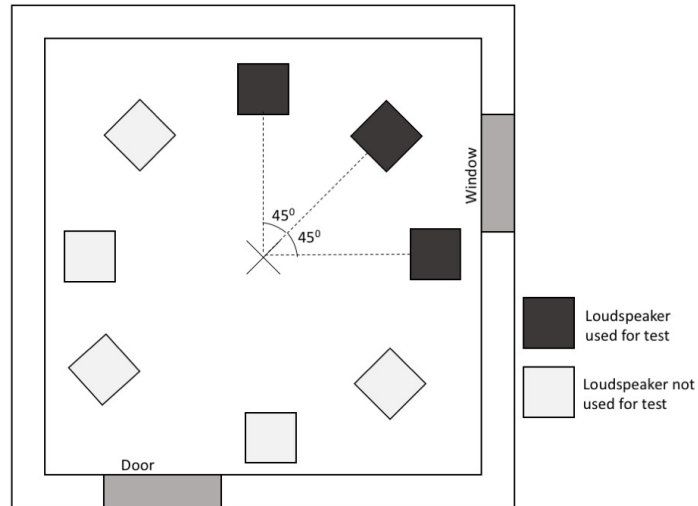


Figure 8: Listening room set-up.

The loudspeakers utilised were Genelec model 8050A loudspeakers. The test script was written in Matlab 2016b and run on a mid-2012 model MacBook Pro. The sound card utilised was a MOTU UltraLite mk3 Hybrid.

### 6.3 Stimuli

A total of three stimuli types were utilised in the tests. The stimulus levels were at an average of 65 dB(A) SPL. Two of the stimuli were pink noise stimuli, generated at a sampling rate of 48 kHz with frequency ranges 20 Hz to 1 kHz and 1 kHz to

8 kHz, respectively. Both noise stimuli were 0.2 s in duration. The third stimuli utilised was speech, in which two-syllable Finnish words are selected at random from a database. The speech stimulus was recorded with a female speaker at a sampling rate of 44.1 kHz and had variable duration.

## 6.4 Test Procedure

The perception of two panning methods was tested: VBAP and time panning. The latter was implemented using the empirical formula in Lee (2017). The listening test was a 2AFC test with six adaptive tracks, one per panning method and stimulus.

A modified adaptive staircase method was utilised for the adaptive tracks; instead of the '2-down 1-up' track a '3-down 1-up' track was implemented in which three trial tests were conducted and the test rendered more difficult if all three trial results were correct. The parameters changed were the panning angle and step size. The initial panning angle was set at 45 degrees; while the step sizes were changed as per the number of reversals, with the initial step size being 10. After the first reversal, the step size changed to 5, then it was changed to 2 after the third reversal. The adaptive tracks were terminated after the reversal count exceeded ten reversals. These step sizes and the termination point were chosen after a series of pilot tests.

At the start of the test, a stimulus, panning method, and direction ('left' or 'right') were chosen at random. If the direction chosen was 'left', the stimulus was first panned to the right of the subject; however, if the direction was 'right', the stimulus was panned to the left. After a 0.1 s pause, the stimulus was then panned to the opposite direction. The subject was then tasked with answering the question "In which direction is the sound going?". The answer was collected utilising a mouse, with subjects pressing the left mouse button for 'left' and the right mouse button for 'right'. This process was then repeated three times, for three trials. In the case of the speech stimulus, two random words were chosen for all three trials and each word utilised once per trial.

If at any point during the test the base panning angle exceeded 45 degrees, it was reset to 45 degrees and the reversal count was increased by 1. Similarly, if the angle decreased below 1 degree, it was set to 1 degree and the reversal count was increased by 1.

Additionally, after every 10 minutes of run time the test was paused in order to give subjects a short break. This was done as the CI subjects experienced fatigue after focusing for extended periods of time. Excluding this break time, the majority of the participants completed the test in 15 minutes.

## 6.5 Calculation of Results

The MAA was calculated as the angle between two stimuli at which the subject's responses were correct for 70% of the trials. Due to the modification of the '2-down 1-up' technique, the calculation of the results had to be modified.

The percentage of accurate responses was calculated per angle, then fitted to a curve. From this curve, it was possible to calculate the MAA. This process was

performed by a Matlab script utilising the built-in libraries for curve fitting. An example of the plots of the percentage of correct response vs the angle, and the fitted curve can be found in [Appendix A](#).

## 6.6 Limitations

As the tests are a 2AFC task, there is the possibility of subjects guessing the right answer. The modified staircase technique was utilised to decrease the chances of this occurring; however, it is still possible that there were some influence from chance answers. Additionally, the levels of the stimuli were not roved from trial to trial. Roving the levels would have reduced the possibility of chance answers.

The second limitation of the study was that there were no listening tests performed with real sound sources; therefore, the MAAs of the CI group for virtual sound sources could not be compared to their MAAs for real sound sources and were instead compared to the results of CI groups in previous studies. However, utilising real sources for the listening test is impractical as it can only be achieved by the positioning numerous loudspeakers as close to one another as possible, which quantises the results, or by an expensive set-up capable of revolving a loudspeaker.

The third limitation is the estimation of the MAA is limited to the range of  $1^\circ$  to  $90^\circ$ , as the base angle between the loudspeakers is  $90^\circ$ . Therefore, for subjects with an MAA larger than or equal to  $90^\circ$ , their calculated MAA would be  $90^\circ$ ; while for subjects with an MAA smaller than or equal to  $1^\circ$ , their calculated MAA would be  $1^\circ$ .

## 7 Results and Discussion

This chapter contains tables and figures of the results of the listening tests and a discussion of them. The MAAs are listed for the CI group and the NH group, and scatter plots are utilised to visualise the correlation between the CI subjects' data and their results.

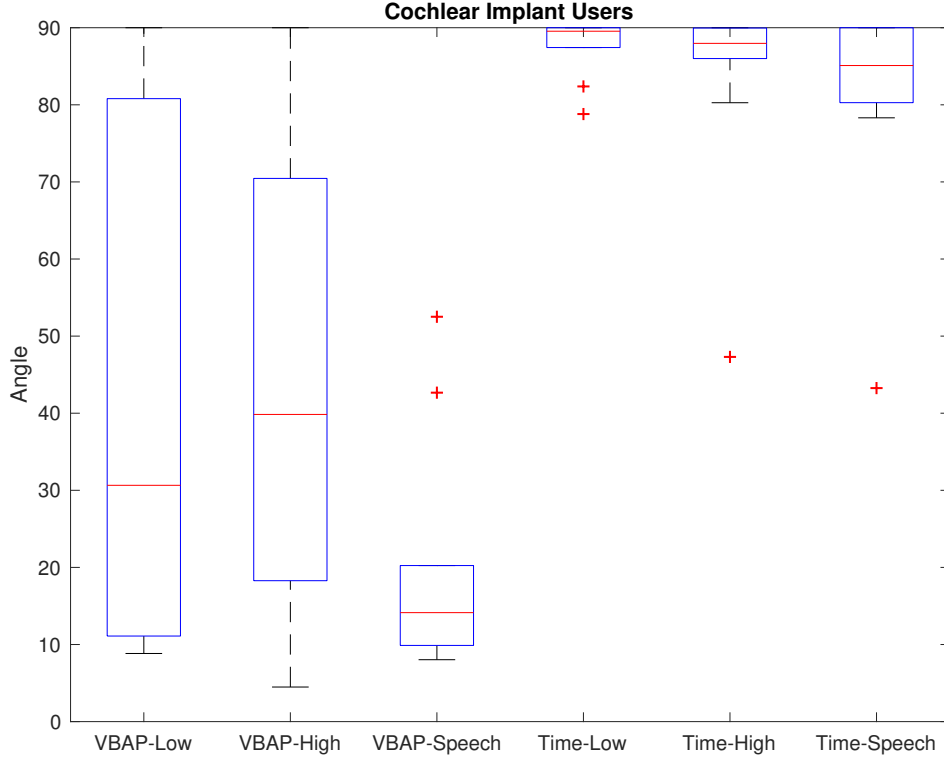


Figure 9: CI group: Box plot of results for VBAP and time panned speech stimuli with a sampling rate of 44.1 kHz, low frequency pink noise stimuli with a sampling rate of 48 kHz subjected to a 400 order FIR bandpass filter with cut-offs 20 Hz and 1 kHz, and high frequency pink noise stimuli with a sampling rate of 48 kHz subjected to a 400 order FIR bandpass filter with cut-offs 1 kHz and 8 kHz.

Figure 9 is a box plot of the MAAs calculated for the CI subjects, and Figure 10 is the corresponding box plot for the NH subjects. Tables 4 and 5 list the MAAs for each subject per panning method and stimuli for the CI group and NH group respectively.

Figure 9 and Table 4 show that the CI subjects were capable of perceiving the VBAP panned sources, with an average MAAs of  $19.96^\circ$  for the speech stimuli,  $40.83^\circ$  for the low frequency noise stimulus, and  $43.28^\circ$  for the high frequency noise stimulus. The NH group achieved MAAs of  $5.72^\circ$  for the speech stimuli,  $4.98^\circ$  for the low frequency noise stimulus, and  $7.54^\circ$  for the high frequency noise stimulus. The CI

Sub	VBAP Low Freq	VBAP High Freq	VBAP Speech	Time Low Freq	Time High Freq	Time Speech
CI1	11.10°	4.48°	12.40°	90.00°	90.00°	87.36°
CI2	46.00°	46.57°	9.88°	90.00°	86.00°	86.00°
CI3	38.53°	18.44°	8.02°	82.38°	80.27°	78.31°
CI4	11.56°	9.78°	9.83°	87.43°	47.20°	80.28°
CI5	10.47°	18.28°	12.32°	78.80°	89.96°	80.33°
CI6	80.80°	70.45°	20.24°	89.17°	90.00°	43.26°
CI7	8.83°	51.72°	15.87°	89.92°	87.27°	90.00°
CI8	88.28°	90.00°	52.51°	90.00°	86.60°	90.00°
CI9	90.00°	90.00°	42.66°	90.00°	88.57°	90.00°
CI10	22.76°	33.11°	15.89°	89.07°	89.24°	84.19°
<b>Mean</b>	40.83°	43.28°	19.96°	87.68°	83.53°	80.97°
<b>Std. Dev.</b>	33.85°	31.90°	15.17°	3.91°	13.07°	13.95°

Table 4: CI Group: Results.

Sub	VBAP Low Freq	VBAP High Freq	VBAP Speech	Time Low Freq	Time High Freq	Time Speech
NH1	3.67°	6.46°	3.75°	26.36°	4.20°	6.74°
NH2	5.04°	13.65°	10.23°	20.28°	11.38°	5.10°
NH3	4.17°	2.00°	4.91°	10.23°	8.00°	8.23°
NH4	3.00°	4.79°	4.68°	7.41°	4.00°	4.00°
NH5	2.91°	6.40°	3.30°	6.36°	5.68°	5.74°
NH6	12.58°	20.30°	8.83°	25.66°	36.79°	35.55°
NH7	5.29°	3.88°	5.33°	8.82°	7.29°	4.76°
NH8	3.30°	4.11°	2.40°	4.20°	4.20°	3.90°
NH9	6.98°	5.13°	7.36°	16.51°	8.30°	9.07°
NH10	2.91°	8.70°	6.38°	4.20°	7.81°	5.52°
<b>Mean</b>	4.98°	7.54°	5.72°	13.00°	9.76°	8.86°
<b>Std. Dev.</b>	2.97°	5.50°	2.49°	8.56°	9.77°	9.53°

Table 5: NH Group: Results.

users' average MAAs are larger than the NH groups', mirroring the trend found in localisation studies with real sound sources.

The CI subjects largely displayed an inability to localise the time panned sources, particularly in comparison to the control group subjects. The best estimate of the average MAA was in the localisation of the time panned speech, 80.98°, which is significantly higher than the 8.86° achieved by the NH group. The majority of the CI subjects were completely unable to localise the time panned stimuli. It may be noted, however, that CI3 and CI6 displayed a greater than chance success in localisation.

The ability of CI subjects to localise the amplitude panned sources depends

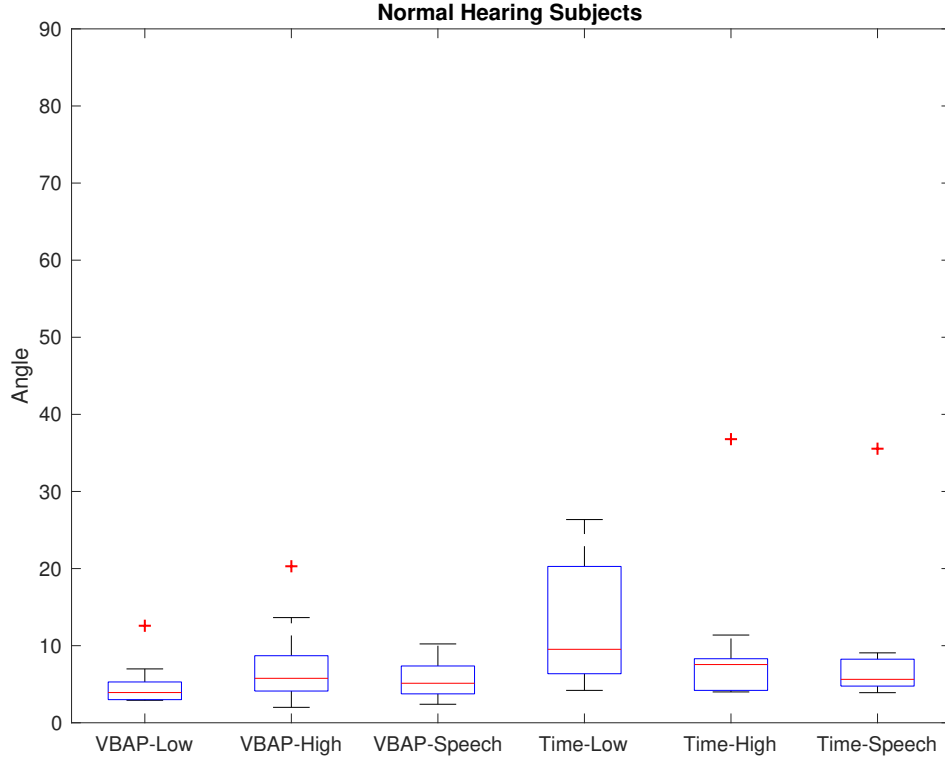


Figure 10: NH group: Box plot of results for VBAP and time panned speech stimuli with a sampling rate of 44.1 kHz, low frequency pink noise stimuli with a sampling rate of 48 kHz subjected to a 400 order FIR bandpass filter with cut-offs 20 Hz and 1 kHz, and high frequency pink noise stimuli with a sampling rate of 48 kHz subjected to a 400 order FIR bandpass filter with cut-offs 1 kHz and 8 kHz.

primarily on the ILD cues, as their ITD cue is severely impoverished. The results of the listening tests must, therefore, be analysed based solely on the ILD cue. In the case of amplitude panned sources, the ILD cues are stable, reliable, and coincide with the values for real sources, as discussed in 4.1.1. Therefore, it follows that the CI subjects can localise amplitude panned virtual sources adequately in the desired direction and have relatively good test results, which are similar to their localisation of real sound sources, which was described in 5.2.1. In the case of time panned sources, the ILD cue is misleading, as discussed in 4.2. The CI subjects' heavy reliance on this cue presumably lead them to perceive the virtual sound source to be in the opposite direction to the desired direction; therefore resulting in poor results for the time panned sources.

A few subjects repeated the test, and displayed an improvement in their MAA for time panned stimuli, a trend that was also noted in Litovsky et al. (2006). This indicates a possibility of training CI users to improve their localisation skills. This is something that has not been attempted before and would be an interesting topic for



future research.

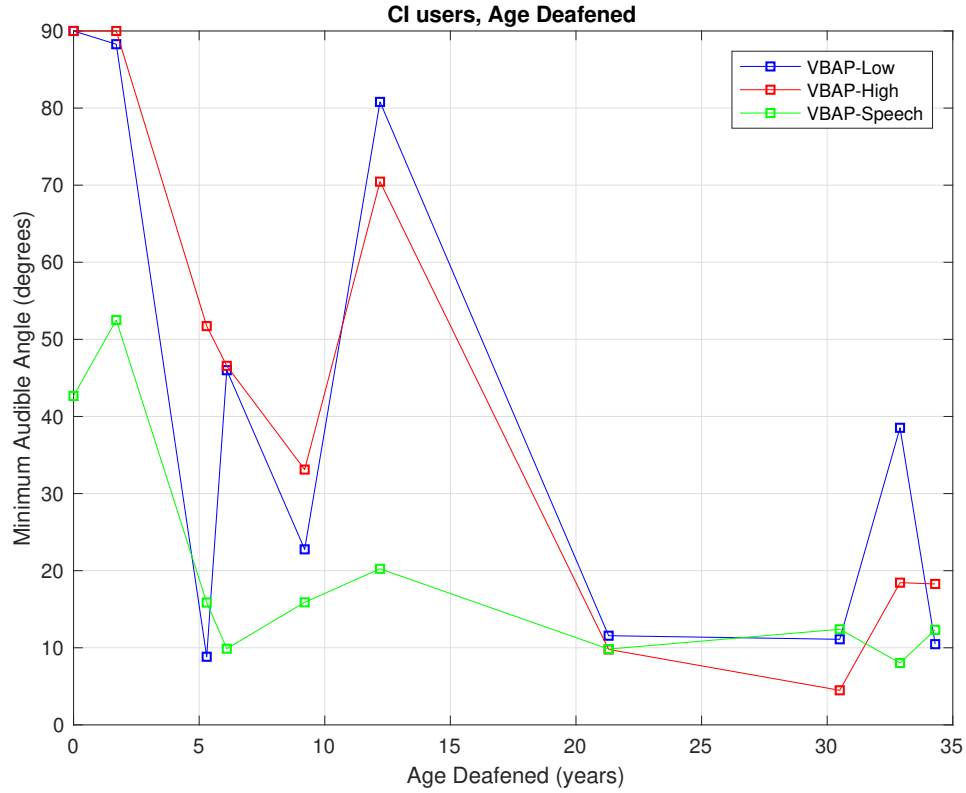


Figure 11: Comparison between the age at which CI subjects were diagnosed deaf and the MAAs results for VBAP panned speech stimuli with a sampling rate of 44.1 kHz, low frequency pink noise stimuli with a sampling rate of 48 kHz subjected to a 400 order FIR bandpass filter with cut-offs 20 Hz and 1 kHz, and high frequency pink noise stimuli with a sampling rate of 48 kHz subjected to a 400 order FIR bandpass filter with cut-offs 1 kHz and 8 kHz.

Figure 9 and Table 4 show large standard deviations of  $33.85^\circ$  and  $31.90^\circ$  in the MAA for the VBAP noise stimuli; contrarily, the standard deviation for the VBAP speech stimuli,  $15.17^\circ$  is smaller by a factor greater than 2. This, in addition to the MAA results, indicates that the duration of the noise stimuli may have presented difficulties for some of the CI subjects, particularly as those incapable of localising the noise stimuli fared better in localising the speech stimuli. Additionally, the difference in the localisation of the speech stimuli MAA may have been due to the fact their CI devices were optimised for speech, or the result of the general trend of CI users finding speech easier to localise than noise. Nonetheless, it may be beneficial to repeat the tests with longer noise stimuli, particularly as the previous studies in CI users' sound localisation utilised noise stimuli that were longer in duration; for example, Majdak et al. (2011) had stimuli of 500 ms, Senn et al. (2005) had stimuli of 1000 ms, and Sechler et al. (2017) had 5-second-long looped stimuli.

While the CI subjects did exhibit a poorer performance in general when compared to the control group subjects, some of them did perform better than the worst performing control group subject. Their mean MAA for amplitude panned speech stimuli was under  $20^\circ$ . These localisation results match the results of CI subjects in MAA tests utilising real sound sources (Litovsky et al., 2006; Senn et al., 2005). This suggests that amplitude panning is effective in producing virtual sound sources for CI subjects. Therefore, it may be utilised as an additional tool in testing their abilities. Furthermore, it suggests that other virtual sound reproduction methods based on amplitude panning may be utilised for testing or to recreate sound scenes.

A correlation was sought between the patient demographic data and the results of the tests. The VBAP MAAs were compared against various factors, such as the age at the time of taking the test, the age of the subject when they received their first implant, the age of the subject when they received the second implant, the time between implants, and the time since the second implantation took place.

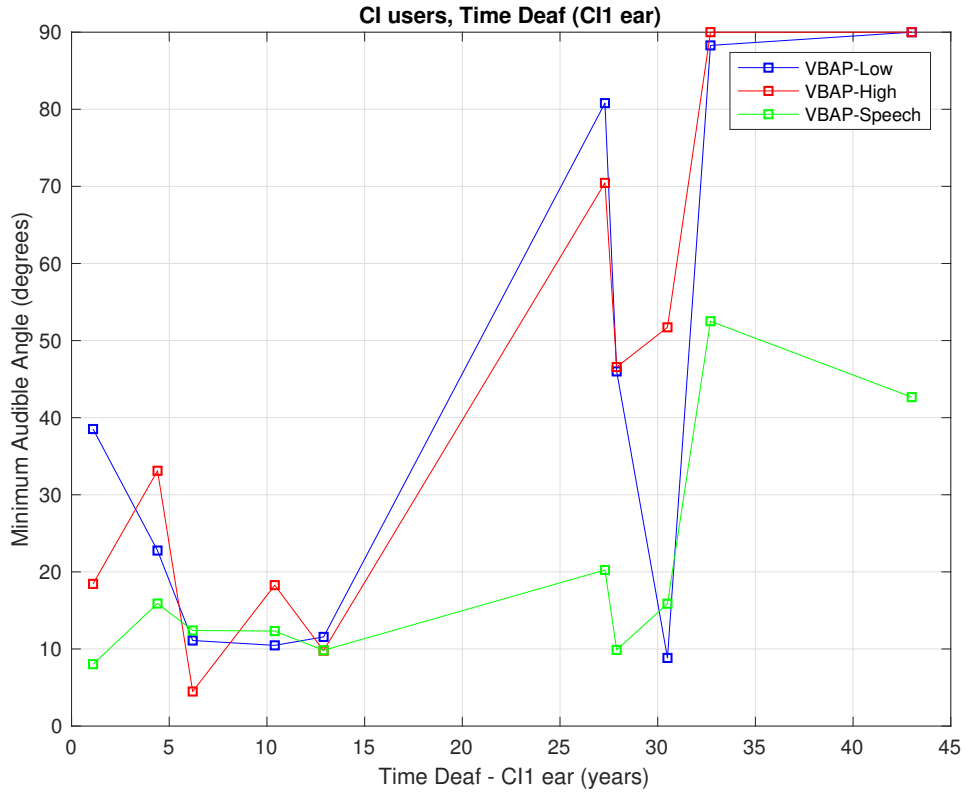


Figure 12: Comparison between the time between diagnoses and ear 1 receiving an implant and the MAAs results for VBAP panned speech stimuli with a sampling rate of 44.1 kHz, low frequency pink noise stimuli with a sampling rate of 48 kHz subjected to a 400 order FIR bandpass filter with cut-offs 20 Hz and 1 kHz, and high frequency pink noise stimuli with a sampling rate of 48 kHz subjected to a 400 order FIR bandpass filter with cut-offs 1 kHz and 8 kHz.

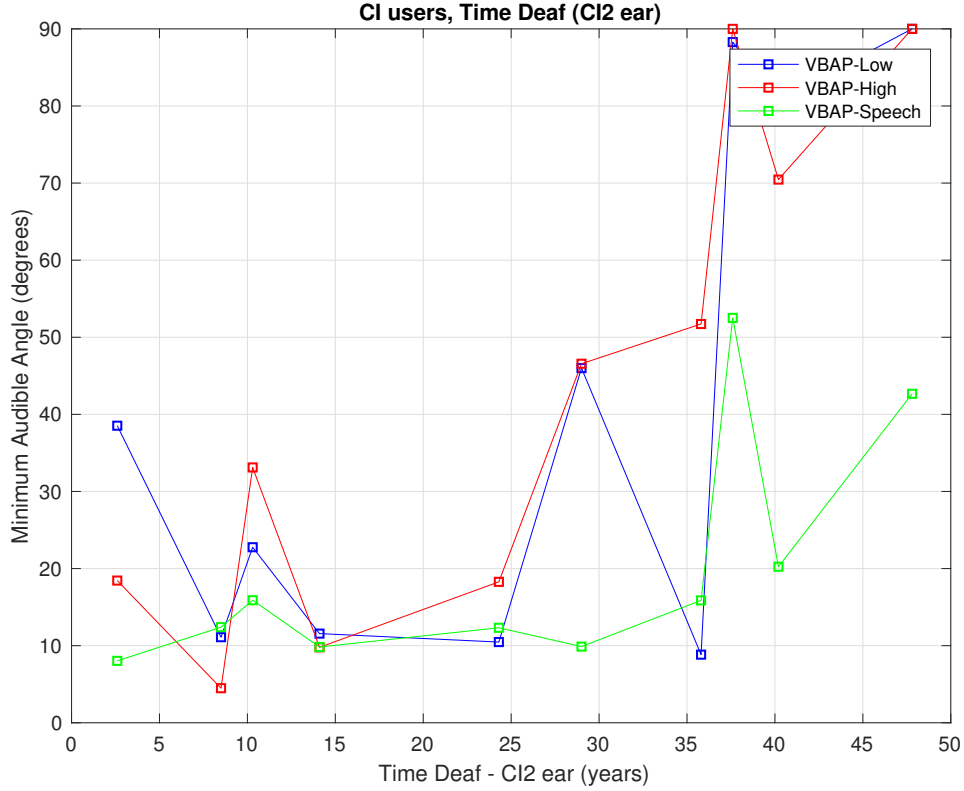


Figure 13: Comparison between the time between diagnoses and ear 2 receiving an implant and the MAAs results for VBAP panned speech stimuli with a sampling rate of 44.1 kHz, low frequency pink noise stimuli with a sampling rate of 48 kHz subjected to a 400 order FIR bandpass filter with cut-offs 20 Hz and 1 kHz, and high frequency pink noise stimuli with a sampling rate of 48 kHz subjected to a 400 order FIR bandpass filter with cut-offs 1 kHz and 8 kHz.

There appears to be a trend between the CI subjects and the age at which they were deafened. Those diagnosed as deaf at a young age, particularly pre-lingually, had large MAAs values, especially for localisation of the noise stimuli. In Figure 11, Subjects CI9 and CI8 were both deafened pre-lingually and are clear examples of this. Figures 12 and 13 show that the time between diagnoses of deafness and the surgery for implantation also influences the results. It would appear that the longer a subject was deaf before the implant, the more difficulty they had in localisation, particularly of the noise stimuli. Furthermore, in Figure 11, the difference between the MAAs for noise stimuli and the speech stimulus appears to be partially affected by the age of diagnoses. This can also be seen in Figures 12 and 13, where the longer the time period between diagnoses and surgery, the more likely it is for there to be a large difference between the MAA for the speech stimulus and for the noise stimuli. It is possible that the difference in standard deviations for the CI group and means may have been due to these factors, and not the duration of the stimuli or

the processing of their devices.

Apart from the above trends, no single variable dependency was found. A possible avenue for future research would be to perform a multivariate analysis of these data.

## 8 Conclusion

The aim of this thesis was to investigate CI users' localisation ability of virtual sound sources. There were two goals. First, to study the utility of virtual sound sources as a tool to test the CI users, as utilising real sources in tests can be difficult and impractical. Secondly, it was hoped that a connection would be found between the listening test results and the CI subjects' data, which could reveal information about user adaptation to the CI devices.

First, it was discovered that the CI users who participated in the listening test displayed a significantly poorer perception of time panned sound sources; on the other hand, their localisation of the stimuli panned utilising VBAP was similar to CI users' localisation of real sound sources, which were studied in Litovsky et al. (2006) and Senn et al. (2005). The average MAA of the CI subjects was found to be  $40.83^\circ$  and  $43.28^\circ$  for the VBAP panned noise stimuli, while their mean MAA for the VBAP panned speech stimuli was under  $19.96^\circ$ . Due to the highly diminished ITD cue received by CI users, their ability to localise amplitude panned sources can be attributed to the stable ILD cue present in the time panned sound sources, while their inability to localise time panned sources can be attributed to them being confused by the unstable ILD cue of the time panned sound sources. Additionally, two of the CI subjects did display weak localisation abilities for the time panned stimuli, indicating that they may receive some localisation information from the envelope ITDs. These results suggest that amplitude panning is perceivable by CI users, and can be utilised as an additional tool for investigating their localisation abilities and perhaps even improve them.

Secondly, the attempts to find a correlation between the listening test results and the patient demographic data of the CI subjects were partially successful. The MAAs derived from the tests were compared with several factors, including the age at time the deaf diagnosis was made and the time between the date of the deaf diagnosis and the date of implantation surgery for each ear. It was found that pre-lingually deafened subjects faced the most difficulty in localisation, particularly for localisation of noise stimuli. The time period between diagnoses and surgery also affected the results, as CI subjects with longer time periods displayed greater localisation difficulties, especially in the case of the noise stimuli.

This research could be further expanded by applying multivariate theory, such as k-clustering, to the data, so as to completely rule out any correlation between patient data and their performance. Furthermore, it could be worthwhile to repeat the experiments with longer noise stimuli to ensure the difference in localisation of noise stimuli and speech stimuli was not caused by the short duration of the noise stimuli. Another avenue of future research would be to repeat the tests with the same subjects and real sound sources, so as to truly compare the MAAs for virtual sound sources to the MAAs for real sound sources.

## References

- Alberti, P. W. (2001). The anatomy and physiology of the ear and hearing, *Occupational exposure to noise: Evaluation, prevention, and control* pp. 53–62.
- Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*, MIT press.
- Buell, T. N. and Hafter, E. R. (1991). Combination of binaural information across frequency bands, *The Journal of the Acoustical Society of America* **90**(4): 1894–1900.
- Clark, G. (2006). *Cochlear implants: fundamentals and applications*, Springer Science & Business Media.
- Cochlear (2017a). Cochlear nucleus ci24re contour advance. Accessed: 2017-11-04.  
**URL:** <http://tinyurl.com/yaw87fn8>
- Cochlear (2017b). Cochlear nucleus ci422 with slim straight. Accessed: 2017-11-04.  
**URL:** <http://tinyurl.com/y7e5ascr>
- Davis, A. C. (1989). The prevalence of hearing impairment and reported hearing disability among adults in great britain, *International Journal of Epidemiology* **18**(4): 911–917.
- Davis, H. and Silverman, S. R. (1970). *Hearing and deafness*, Holt, Rinehart & Winston of Canada Ltd.
- Dorman, M. F., Loisel, L. H., Cook, S. J., Yost, W. A. and Gifford, R. H. (2016). Sound source localization by normal-hearing listeners, hearing-impaired listeners and cochlear implant listeners, *Audiology and Neurotology* **21**(3): 127–131.
- Dorman, M. F., Loisel, L., Stohl, J., Yost, W. A., Spahr, A., Brown, C. and Cook, S. (2014). Interaural level differences and sound source localization for bilateral cochlear implant patients, *Ear and hearing* **35**(6): 633.
- Duda, R. O. and Martens, W. L. (1998). Range dependence of the response of a spherical head model, *The Journal of the Acoustical Society of America* **104**(5): 3048–3058.
- Dunn, C. C., Tyler, R. S., Oakley, S., Gantz, B. J. and Noble, W. (2008). Comparison of speech recognition and localization performance in bilateral and unilateral cochlear implant users matched on duration of deafness and age at implantation, *Ear and Hearing* **29**(3): 352.
- Fishman, K. E., Shannon, R. V. and Slattery, W. H. (1997). Speech recognition as a function of the number of electrodes used in the speak cochlear implant speech processor, *Journal of Speech, Language, and Hearing Research* **40**(5): 1201–1215.

- Gold, B., Morgan, N. and Ellis, D. (2011). *Speech and audio signal processing: processing and perception of speech and music*, John Wiley & Sons.
- Grantham, D. W., Ashmead, D. H., Ricketts, T. A., Labadie, R. F. and Haynes, D. S. (2007). Horizontal-plane localization of noise and speech signals by postlingually deafened adults fitted with bilateral cochlear implants, *Ear and Hearing* **28**(4): 524–541.
- Grieco-Calub, T. M. and Litovsky, R. Y. (2012). Spatial acuity in two-to-three-year-old children with normal acoustic hearing, unilateral cochlear implants and bilateral cochlear implants, *Ear and hearing* **33**(5): 561.
- Härkönen, K., Kivekäs, I., Rautiainen, M., Kotti, V., Sivonen, V. and Vasama, J.-P. (2015). Sequential bilateral cochlear implantation improves working performance, quality of life, and quality of hearing, *Acta oto-laryngologica* **135**(5): 440–446.
- Kerber, I. S. and Seeber, I. B. U. (2012). Sound localization in noise by normal-hearing listeners and cochlear implant users, *Ear and hearing* **33**(4): 445.
- Laszig, R., Aschendorff, A., Stecker, M., Muller-Deile, J., Maune, S., Dillier, N., Weber, B., Hey, M., Begall, K., Lenarz, T. et al. (2004). Benefits of bilateral electrical stimulation with the nucleus cochlear implant in adults: 6-month postoperative results, *Otology & Neurotology* **25**(6): 958–968.
- Lawson, D. T., Wilson, B. S., Zerbi, M., Van Den Honert, C., Finley, C. C., Farmer Jr, J. C., McElveen Jr, J. T. and Roush, P. A. (1998). Bilateral cochlear implants controlled by a single speech processor., *Otology & Neurotology* **19**(6): 758–761.
- Lee, H. (2017). Perceptually motivated amplitude panning (pmap) for accurate phantom image localization, *Audio Engineering Society Convention 142*, Audio Engineering Society.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics, *The Journal of the Acoustical society of America* **49**(2B): 467–477.
- Lipshitz, S. P. (1986). Stereo microphone techniques: Are the purists wrong?, *Journal of the Audio Engineering Society* **34**(9): 716–744.
- Litovsky, R. Y., Colburn, H. S., Yost, W. A. and Guzman, S. J. (1999). The precedence effect, *The Journal of the Acoustical Society of America* **106**(4): 1633–1654.
- Litovsky, R. Y., Johnstone, P. M., Godar, S., Agrawal, S., Parkinson, A., Peters, R. and Lake, J. (2006). Bilateral cochlear implants in children: localization acuity measured with minimum audible angle, *Ear and hearing* **27**(1): 43.
- Litovsky, R. Y., Parkinson, A. and Arcaroli, J. (2009). Spatial hearing and speech intelligibility in bilateral cochlear implant users, *Ear and hearing* **30**(4): 419.

- Litovsky, R. Y., Parkinson, A., Arcaroli, J., Peters, R., Lake, J., Johnstone, P. and Yu, G. (2004). Bilateral cochlear implants in adults and children, *Archives of Otolaryngology-Head & Neck Surgery* **130**(5): 648–655.
- Loizou, P. C. (1999). Introduction to cochlear implants, *IEEE Engineering in Medicine and Biology Magazine* **18**(1): 32–42.
- Majdak, P., Goupell, M. J. and Laback, B. (2011). Two-dimensional localization of virtual sound sources in cochlear-implant listeners, *Ear and hearing* **32**(2): 198–208.
- Marieb, E. N. and Hoehn, K. (2007). *Human anatomy & physiology*, Pearson Education.
- Med-El (2017a). Automatic sound management. Accessed: 2017-08-18.  
**URL:** <http://www.medel.com/int/automatic-sound-management/>
- Med-El (2017b). Electrode arrays. Accessed: 2017-11-04.  
**URL:** <http://s3.medel.com/pdf/21617.pdf>
- Med-El (2017c). Rondo audio processor : CI | MED-EL. Accessed: 2017-08-18.  
**URL:** <http://www.medel.com/int/rondo>
- Med-El (2017d). Sonnet audio processor : CI | MED-EL. Accessed: 2017-08-18.  
**URL:** <http://www.medel.com/int/sonnet>
- Moller, A. R. (2012). *Hearing: anatomy, physiology, and disorders of the auditory system*, Plural Publishing.
- Møller, H., Sørensen, M. F., Hammershoi, D. and Jensen, C. B. (1995). Head-related transfer functions of human subjects, *Journal of the Audio Engineering Society* **43**(5): 300–321.
- Moore, B. C. (2012). *An introduction to the psychology of hearing*, Brill.
- Moore, F. R. (1990). *Elements of computer music*, Prentice-Hall, Inc.
- Neuman, A. C., Haravon, A., Sislian, N. and Waltzman, S. B. (2007). Sound-direction identification with bilateral cochlear implants, *Ear and hearing* **28**(1): 73–82.
- Nopp, P., Schleich, P. and D’haese, P. (2004). Sound localization in bilateral users of med-el combi 40/40+ cochlear implants, *Ear and Hearing* **25**(3): 205–214.
- Oh, S.-h., Kim, C.-s., Kang, E. J., Lee, D. S., Lee, H. J., Chang, S. O., Ahn, S.-h., Hwang, C. H., Park, H. J. and Koo, J. W. (2003). Speech perception after cochlear implantation over a 4-year time period, *Acta oto-laryngologica* **123**(2): 148–153.
- Pulkki, V. (1997). Virtual sound source positioning using vector base amplitude panning, *Journal of the Audio Engineering Society* **45**(6): 456–466.



- Pulkki, V. (2002). Microphone techniques and directional quality of sound reproduction, *Audio Engineering Society Convention 112*, Audio Engineering Society.
- Pulkki, V. and Karjalainen, M. (2015). *Communication acoustics: an introduction to speech, audio and psychoacoustics*, John Wiley & Sons.
- Pulkki, V., Karjalainen, M. and Huopaniemi, J. (1999). Analyzing virtual sound source attributes using a binaural auditory model, *Journal of the Audio Engineering Society* **47**(4): 203–217.
- Sechler, S. D., Valdes, A. L., Waechter, S. M., Simoes-Franklin, C., Viani, L. and Reilly, R. B. (2017). Virtual reality sound localization testing in cochlear implant users, *Neural Engineering (NER), 2017 8th International IEEE/EMBS Conference on*, IEEE, pp. 379–382.
- Seeber, B. and Fastl, H. (2004). Localization cues with bilateral cochlear implants investigated in virtual space—a case study, *Proc Joint Congress CFA/DAGA*, Vol. 4, pp. 213–214.
- Seeber, B. U. and Fastl, H. (2008). Localization cues with bilateral cochlear implants, *The Journal of the Acoustical Society of America* **123**(2): 1030–1042.
- Senn, P., Kompis, M., Vischer, M. and Haeusler, R. (2005). Minimum audible angle, just noticeable interaural differences and speech intelligibility with bilateral cochlear implants using clinical speech processors, *Audiology and Neurotology* **10**(6): 342–352.
- Shaw, E. (1974). Transformation of sound pressure level from the free field to the eardrum in the horizontal plane, *The Journal of the Acoustical Society of America* **56**(6): 1848–1861.
- Smith, A. T. (1997). Visual perception: Physiology, psychology and ecology.
- Tobey, E. A., Geers, A. E., Douek, B. M., Perrin, J., Skellett, R., Brenner, C. and Toretta, G. (2000). Factors associated with speech intelligibility in children with cochlear implants, *Annals of Otology, Rhinology & Laryngology* **109**(12\_suppl): 28–30.
- Tyler, R. S., Noble, W., Dunn, C. and Witt, S. (2006). Some benefits and limitations of binaural cochlear implants and our ability to measure them: Algunos beneficios y limitaciones de los implantes cocleares binaurales y nuestra capacidad para medirlos, *International Journal of Audiology* **45**(sup1): 113–119.
- Välimaa, T., Kunnari, S., Laukkanen-Nevala, P. and Lonka, E. (n.d.). Early vocabulary development in children with bilateral cochlear implants, *International Journal of Language & Communication Disorders*.
- Välimaa, T. T., Sorri, M. J., Laitakari, J., Sivonen, V. and Muhli, A. (2011). Vowel confusion patterns in adults during initial 4 years of implant use, *Clinical linguistics & phonetics* **25**(2): 121–144.

- van de Par, S. and Kohlrausch, A. (1997). A new approach to comparing binaural masking level differences at low and high frequencies, *The Journal of the Acoustical Society of America* **101**(3): 1671–1680.
- Van Hoesel, R. and Clark, G. (1997). Psychophysical studies with two binaural cochlear implant subjects, *The Journal of the Acoustical Society of America* **102**(1): 495–507.
- van Hoesel, R. J. and Tyler, R. S. (2003). Speech perception, localization, and lateralization with bilateral cochlear implants, *The Journal of the Acoustical Society of America* **113**(3): 1617–1630.
- Van Hoesel, R., Tong, Y., Hollow, R. and Clark, G. (1993). Psychophysical and speech perception studies: a case report on a binaural cochlear implant subject, *The Journal of the Acoustical Society of America* **94**(6): 3178–3189.
- WHO (2012). Who global estimates on prevalence of hearing loss, *Geneva: World Health Organization* .
- Yost, W. A. and Dye Jr, R. H. (1988). Discrimination of interaural differences of level as a function of frequency, *The Journal of the Acoustical Society of America* **83**(5): 1846–1851.
- Zheng, Y., Godar, S. P. and Litovsky, R. Y. (2015). Development of sound localization strategies in children with bilateral cochlear implants, *PloS one* **10**(8): e0135790.

## A Appendix A

This appendix contains the plots of the calculations of the results of subjects NH4 and CI4, as described in 6.5.

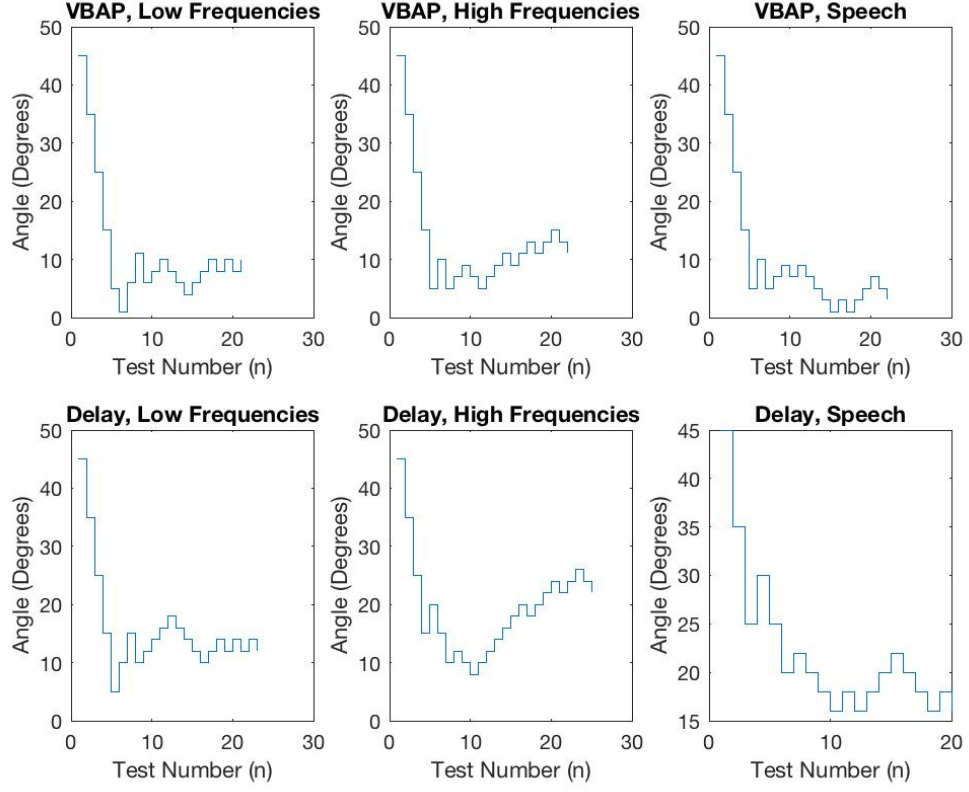


Figure A1: Listening test plot for subject NH4, showing the change in panning angle plotted as a function of test number.

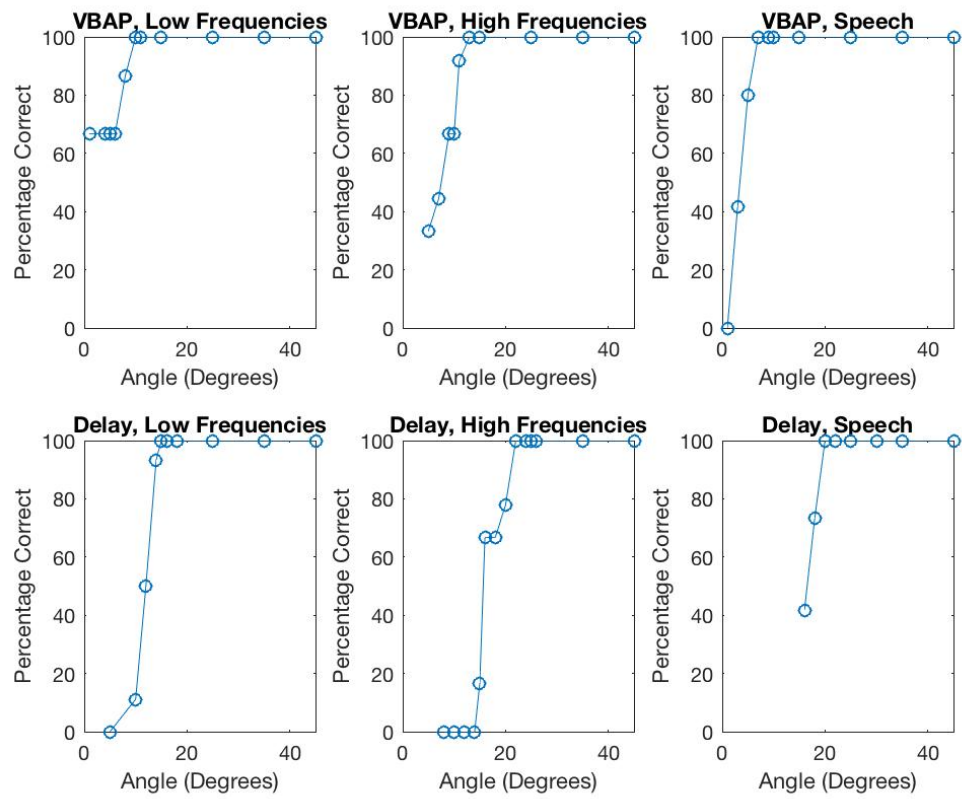


Figure A2: Results for subject NH4 plotted as percentage of correct responses for each panning angle.

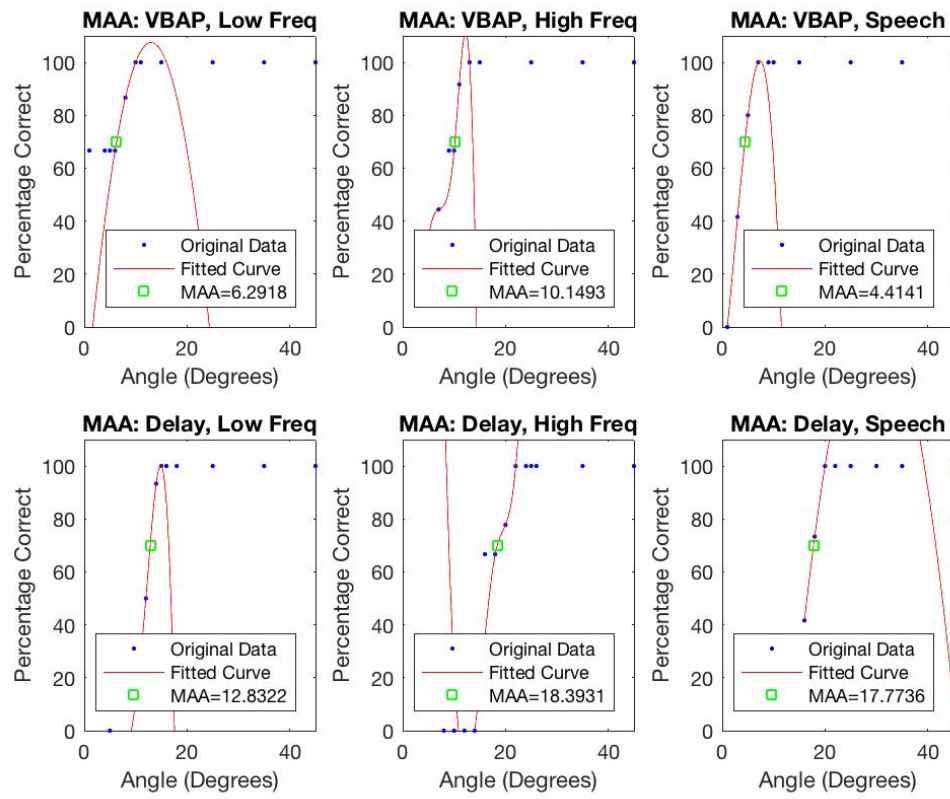


Figure A3: Curve fitted results for subject NH4, with the MAA marked.

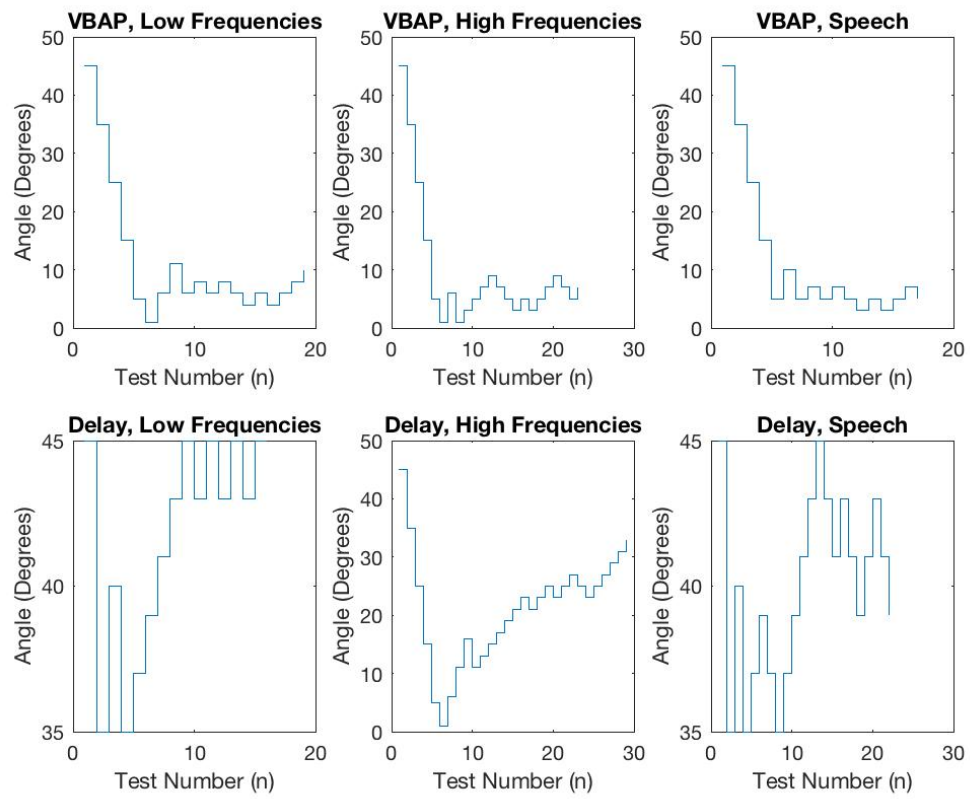


Figure A4: Listening test plot for subject CI4, showing the change in panning angle plotted as a function of test number.

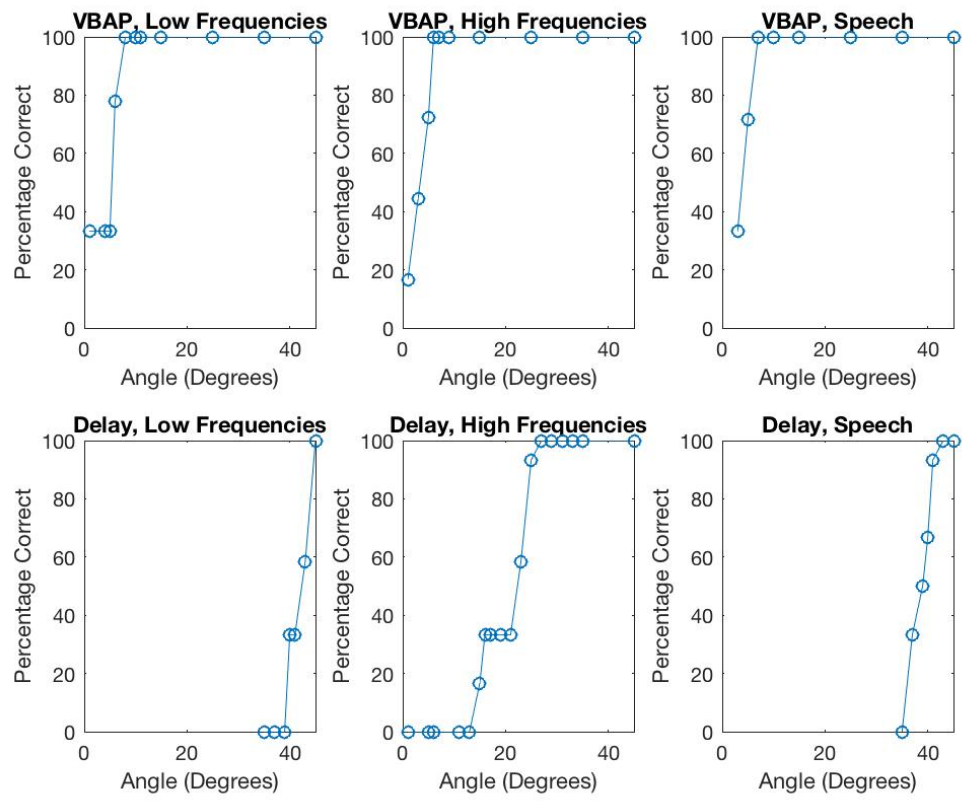


Figure A5: Results for subject CI4 plotted as percentage of correct responses for each panning angle.

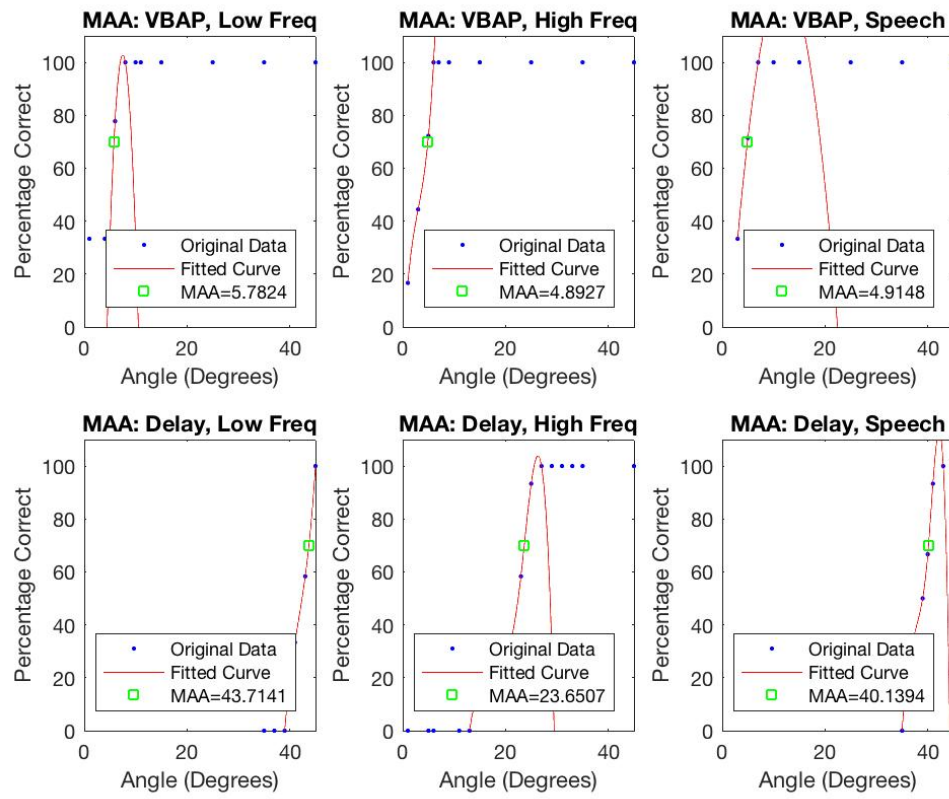


Figure A6: Curve fitted results for subject CI4, with the MAA marked.